Lockheed Martin Center for Innovation

MORS

DoE Tutorial

75TH MORSS
US Naval Academy
Annapolis, Maryland
June 12-13-14 2007
Through Multiple Disciplines, Analytical Power

712CD

75TH MORSS CD Cover Page

If you would like your presentation included in the 75th MORSS Final Report CD it must :

1. *Be unclassified, approved for public release, distribution unlimited, and is exempt from U.S. export licensing and other export approvals including the International Traffic in Arms Regulations (22CFR120 et seq.);*
2. Include MORS Form 712CD as the first page of the presentation;
3. Have an approved MORS form 712 A/B and
4. Be turned into the MORS office no later than: **DEADLINE: 14 June 2007 (Late submissions will not be included.)**

Author Request (To be completed by applicant) - The following author(s) request authority to disclose the following presentation in the MORSS Final Report, for inclusion on the MORSS CD and/or posting on the MORS web site.

Name of Principal Author and all other author(s):

Paul J. Bross

Principal Author's Organization and address:	
Lockheed Martin Corporation Center for Innovation 7021 Harbour View Blvd., Suite 105 Suffolk, VA 23435	Phone: 757-935-9504 Fax: 757-935-9563 Email: paul.bross@lmco.com

Please use the same title listed on the 75TH MORSS Disclosure Form 712 A/B. If the title of the presentation has changed please list both.)

Original title on 712 A/B:

Design of Experiments: A Tutorial

If the title was revised please list the original title above and the revised title here:

PRESENTED IN:

WORKING GROUP:	33	DEMONSTRATION:	
COMPOSITE GROUP:		POSTER:	
SPECIAL SESSION 1:		TUTORIAL:	
SPECIAL SESSION 2:		OTHER:	
SPECIAL SESSION 3:			

This presentation is believed to be: *Unclassified, approved for public release, distribution unlimited, and is exempt from U.S. export licensing and other export approvals including the International Traffic in Arms Regulations (22CFR120 et seq.)*

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 JUN 2007	**N/A**	**-**

4. TITLE AND SUBTITLE	
Design of Experiments A Tutorial	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Lockheed Martin Corporation Suffolk, VA 22435**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

See also ADM202526. Military Operations Research Society Symposium (75th) Held in Annapolis, Maryland on June 12-14, 2007, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**119**	

Lockheed Martin Center for Innovation

Design of Experiments

A Tutorial

Paul J. Bross
Operations Research Principal
Center for Innovation
June 2007

Tutorial Composition

- Basic Concepts
- Break
- Advanced Concepts
- Break
- Detailed Examples
- Wrap-Up

Design of Experiments

Basic Concepts

Summary Slide

- What is DoE?
- Purposes of Experimenting
- Experimentation Strategies
- Basic Principles
- Nuisance Factors
- Design Steps
- Major Guidelines
- Simple Comparison Experiments
- Single Factor Experiments
- Latin Squares

What is DoE?

- **Experiment**: a test or series of tests where the experimenter makes purposeful changes to input variables of a process or system so that we can observe or identify the reasons for changes in the output responses.
- **Design of Experiments**: is concerned with the planning and conduct of experiments to analyze the resulting data so that we obtain valid and objective conclusions.

Lineage

DoE Tutorial

- 1771 – *Course of Experimental Agriculture*, Arthur Young
  - One of the earliest direct experimental scientific documents
  - Insisted on split-field trials
  - Required repeated trials in different fields
- 1919 – R.A. Fisher started work as a statistician at Rothamsted Agricultural Experimental Station
  - Randomization of trials
  - Creation of the technique "Analysis of Variance"
- Today….

Code of Best Practices (COBP)

- *Code of Best Practice for Experimentation*, CCRP, 2002
- *Campaigns of Experimentation: Pathways to Innovation and Transformation*, Alberts & Hayes, 2005
- These documents identify 3 types of experiments:
  - Discovery
  - Hypothesis
  - Demonstration
- This tutorial focuses on aspects of the first two types

Lockheed Martin Center for Innovation

Types of Experiments

- Discovery
  - Designed to generate new ideas or approaches
  - Usually involve “hands-on” activities
  - May involve systems or processes that are not well understood or refined
- Hypothesis
  - Closer to the traditional academic approach
  - Seek to falsify specific hypotheses
  - Used often in the attempt to “prove” a theory, idea, or approach

Why Experiment?

- Determine which variables are the most influential in a process or system
- Determine where to set the inputs so the output is always near the desired state
- Determine where to set the inputs so the output variability is minimized
- Determine where to set the inputs so the influence of uncontrollable factors is minimized (robust design)

Experimentation Strategies

- Best Guess
  - PRO: Works reasonably well when used by SMEs with solid foundational knowledge on known issues
  - CONs:
    - If it fails, need to guess again…and again…until….
    - If get acceptable results first time, may stop without discovering "better"
- One Factor at a Time
  - PRO: Straight-forward, easily understood
  - CONs:
    - Impossible to address interactions
    - Tends to "over collect" data, not efficient sample sizes
- Factorial
  - PROs:
    - Full evaluation of individual <u>and</u> interaction effects
    - Most efficient design with respect to sample sizes
  - CON: More complex to explain to untrained audiences

Lockheed Martin Center for Innovation

Basic Principles

- Replication
  - Permits estimation of experimental error
  - Permits more precise estimates of the sample statistics
  - Not to be confused with repeated measures
- Randomization
  - Insures that observations or errors are more likely to be independent
  - Helps "average out" effects of extraneous factors
  - Special designs when complete randomization not feasible
- Blocking
  - Designed to improve precision of comparisons
  - Used to reduce or eliminate nuisance factors

Nuisance Factors

- Definition: A <u>nuisance</u> <u>factor</u> is a "design factor that *probably* has an effect on the response but we are not interested in that effect" [Montgomery, p126, emphasis added]
- Nuisance Factors, Types $\Rightarrow$ Cures
  - Known and controllable $\Rightarrow$ Use blocking to systematically eliminate the effect
  - Known but uncontrollable $\Rightarrow$ If it can be measured, use Analysis of Covariance (ANCOVA)
  - Unknown and uncontrollable $\Rightarrow$ Randomization is the insurance

Design Steps

- Recognition and statement of the problem in *nonstatistical* language
- Selection of factors, levels, ranges
- Selection of response variables
- Choice of experimental design
- Performance of the experiment
- Statistical analysis of the data
- Conclusions and recommendations

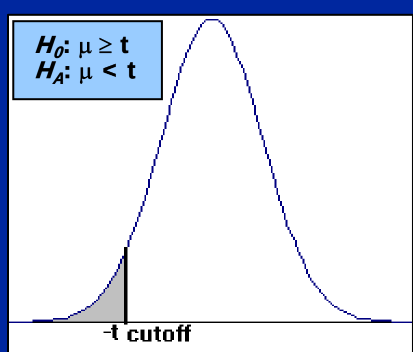
Major Guidelines

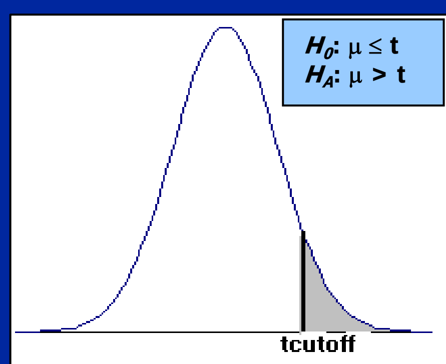
- Use team's non-statistical knowledge of the problem to:
  - Choose factors
  - Determine proper levels
  - Decide number of replications
  - Interpret results
- Keep the design and analysis as simple as possible
- Recognize the difference between practical and statistical significance
- Be prepared to iterate – commit no more than 25% of available resources to first series

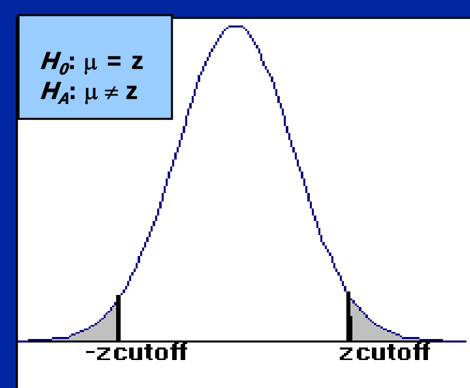
Simple Comparison Experiments

- Goal:
  - Compare two or more means; variances; probabilities
  - Compare *A* versus *B*: [better or worse] – paired comparison is a special case of randomized block design
- Major Considerations
  - Sample size
  - Distributional knowledge: Normal, χ^2, F …..*etc*.
  - Structure of the statistical hypothesis
    - One-tailed

Two-tailed tests







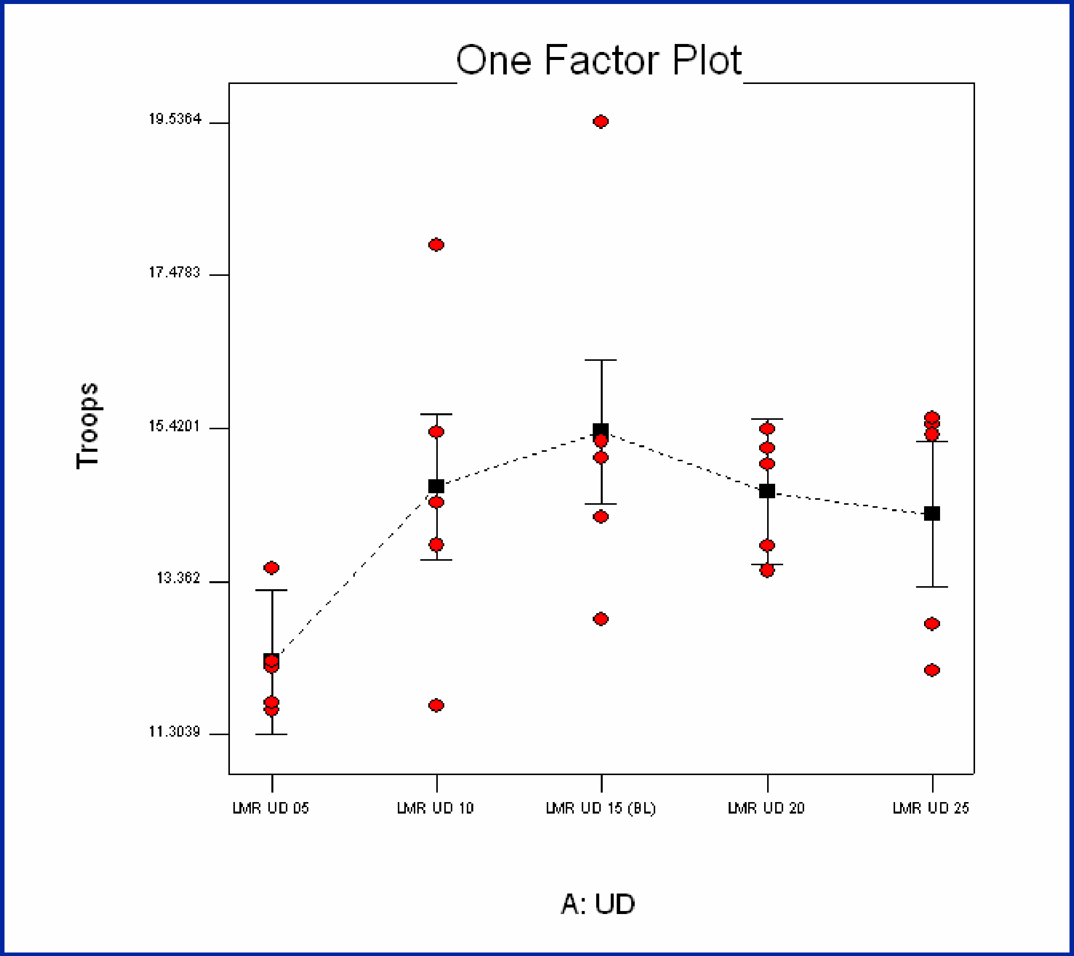
Single Factor Experiments

- One Factor – Multiple Levels
- "One-level-at-a-time" analysis isn't efficient
  - Consider one factor with five levels
  - Pair-wise comparison requires 10 pairs [${}_5C_2 = 10$]
  - If each comparison has $\alpha = 0.05$, then

    Probability(correct assessment) = $(1-\alpha)^{10} = 0.60$
- Technique of Choice – ANOVA
  - Tests hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots \mu_n$
  - Assumptions
    - Error term is Normal $(0,\sigma^2) \Rightarrow$ test residuals to confirm
    - Conditions properly randomized
    - Results are independent; errors are independent
  - If reject H_o *(i.e., failed the test)* then use Newman-Keuls Range Test or Duncan's Multiple Range Test to determine the specifics
  - Note – there are non-parametric tests in lieu of ANOVA if assumptions are not met (*e.g.* Kruskal-Wallis Test)

Single Factor Multiple Levels

DoE Tutorial



- Single Factor – Unit Days of Supply
- Levels – 5, 10, 15, 20, 25

Latin Squares

- Latin Square: An arrangement of conditions such that each combination occurs only once in each row and column of the test matrix.

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

- Graeco-Latin Square: The superposition of two Latin Squares such that each paired-combination occurs only once in each row and column.

Aα	Bδ	Cβ	Dε	Eγ
Bβ	Cε	Dγ	Eα	Aδ
Cγ	Dα	Eδ	Aβ	Bε
Dδ	Eβ	Aε	Bγ	Cα
Eε	Aγ	Bα	Cδ	Dβ

Order 5

Orthogonal Latin Square

Latin Squares – Practical Example

- Conduct a test of new intelligence fusion procedures using four analyst teams examining four scenarios. Each fusion process will take one day of test activity to fully work the process.
  - Day 1 ⇒ Orientation Day for participants; assign teams (A,B,C,D)
  - Days 2 through 5 ⇒ Test days

	Tues	Wed	Thurs	Fri
Scenario 1	A	B	C	D
Scenario 2	B	C	D	A
Scenario 3	C	D	A	B
Scenario 4	D	A	B	C

- Do it again, 3 months later with different teams (α,β,χ,δ)

	Tues	Wed	Thurs	Fri
Scenario 1	α	β	χ	δ
Scenario 2	β	χ	δ	α
Scenario 3	χ	δ	α	β
Scenario 4	δ	α	β	χ

- Combine analytical results

	Day 1	Day 2	Day 3	Day 4
Scenario 1	A α	B β	C χ	D δ
Scenario 2	B β	C χ	D δ	A α
Scenario 3	C χ	D δ	A α	B β
Scenario 4	D δ	A α	B β	C χ

Non-Orthogonal Latin Square

Design of Experiments

Advanced Concepts

Segment Agenda

- Advanced DoE
  - Factorials
    - Full
    - Fractional
    - Other Types
  - Complex Designs

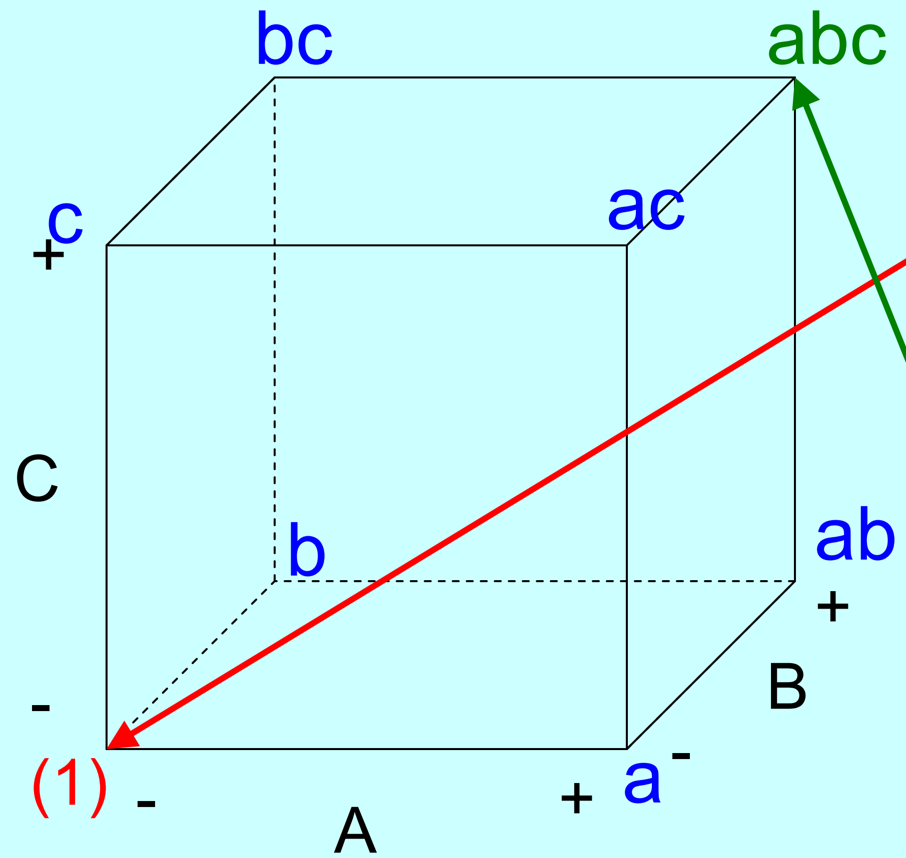
Full Factorial Designs

- Definition: An experiment in which for each completed trial or replication of the experiment all possible combinations of the levels of the factors are investigated.
- Design Notation
  - General Notation for 2-level experiment $\Rightarrow 2^k$ where k = number of factors
    - 3 factors 2 levels each = 2^3 design
  - Factors and Levels $\Rightarrow$ example for 3 factors, 2 levels
    - Aa Bb Cc
    - A^+A^- B^+B^- C^+C^-
    - (1) a b c

Full Factorial Design

- All combinations are examined
  - Example 2^3 design = 8 experimental settings:
    $A^+B^+C^+$ $A^-B^+C^+$ $A^+B^-C^+$ $A^+B^+C^-$ $A^-B^-C^+$ $A^-B^+C^-$ $A^+B^-C^-$ $A^-B^-C^-$
- Effects Evaluated
  - Main effects of single factors: A, B, C
  - Second Order (2-factor) interactions: AB, AC, BC
  - Third Order (3-factor) interactions: ABC
  - In general, a 2^k design evaluates all 1, 2, …k-1, k-factor effects
- Advantages over "one-factor-at-a-time"
  - More efficient in time, resources, sample size
  - Addresses interactions
  - Provides insight over a *range* of experimental conditions

Factorial Efficiency – Graphically (1)

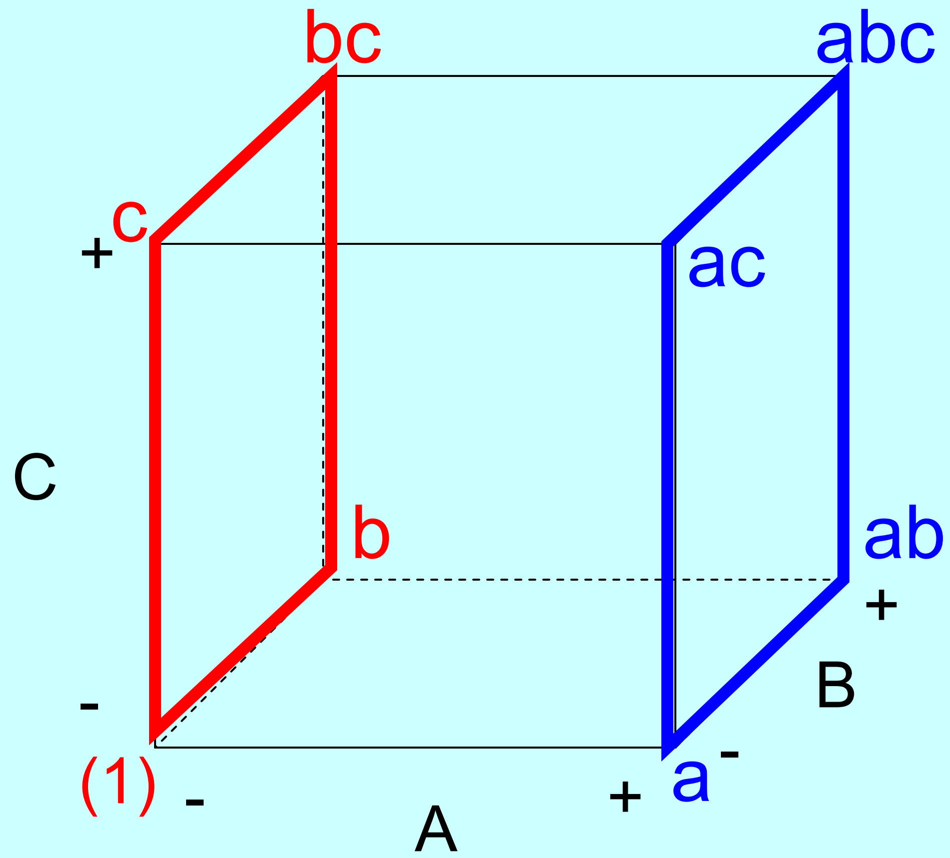


A	B	C
-	-	-
+	-	-
-	+	-
+	+	-
-	-	+
+	-	+
-	+	+
+	+	+

Factorial Efficiency – Graphically (2)

Main Effect A

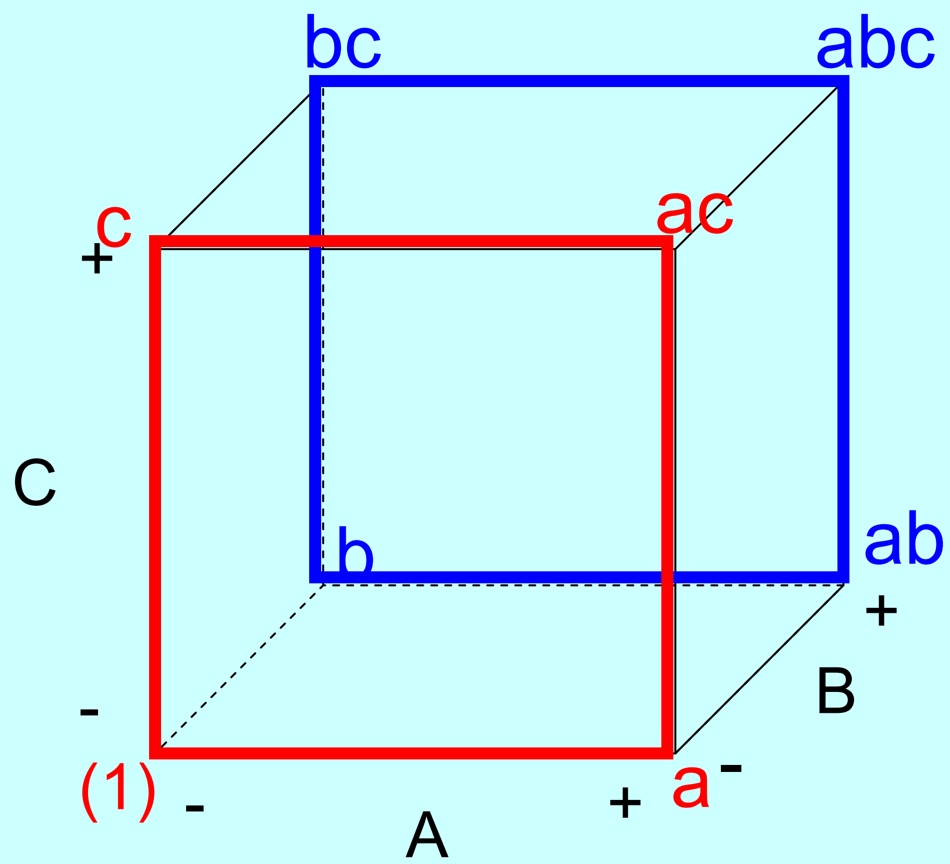
= (1/4*n*) * [blue square
-red square]



Factorial Efficiency – Graphically (3)

Main Effect B

= (1/4*n*) * [blue square
-red square]

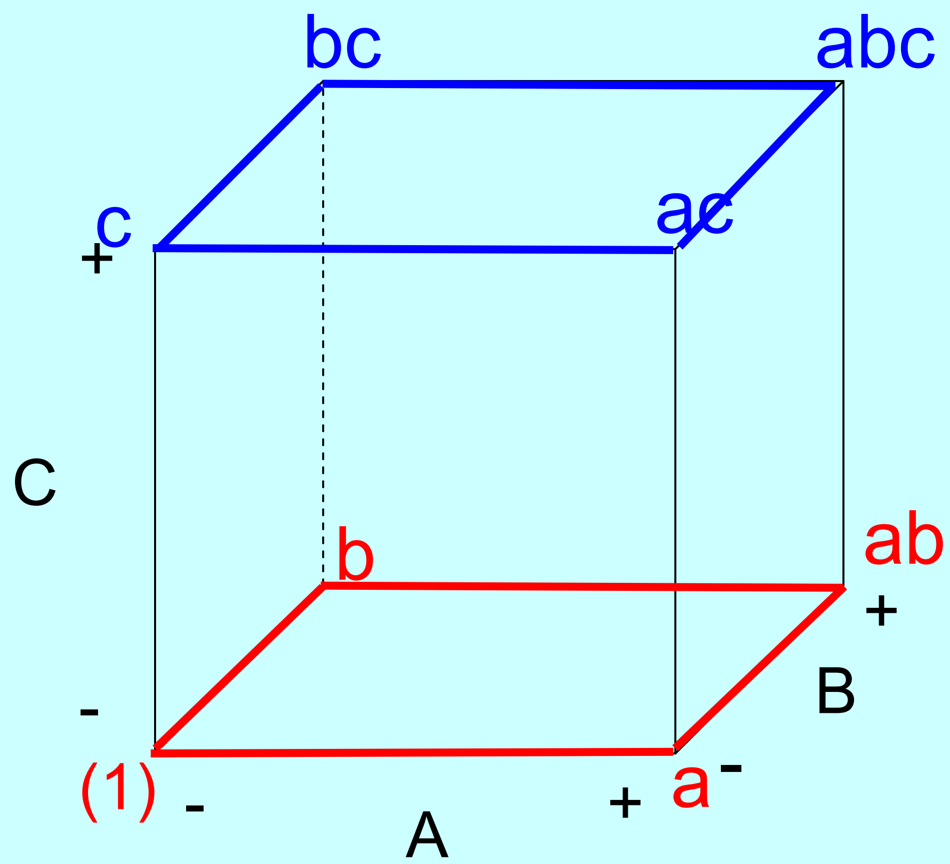


Factorial Efficiency – Graphically (4)

Main Effect C

= (1/4*n*) * [blue square

-red square]

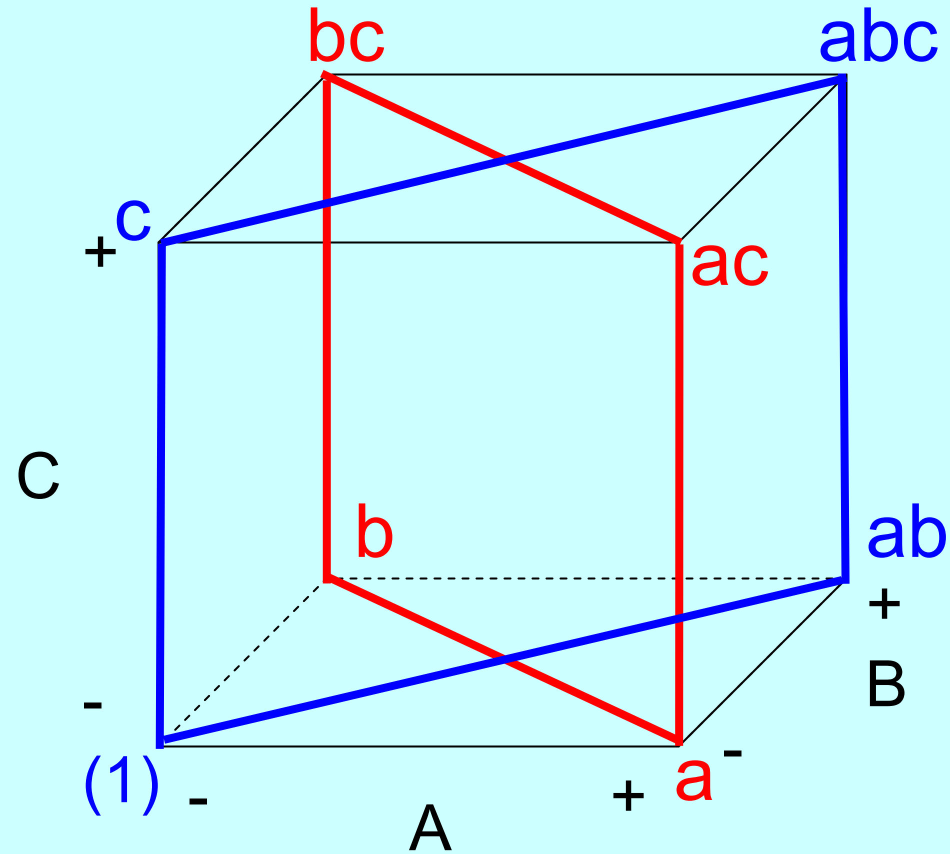


Factorial Efficiency – Graphically (5)

Effect AB

= (1/4*n*) * [blue plane

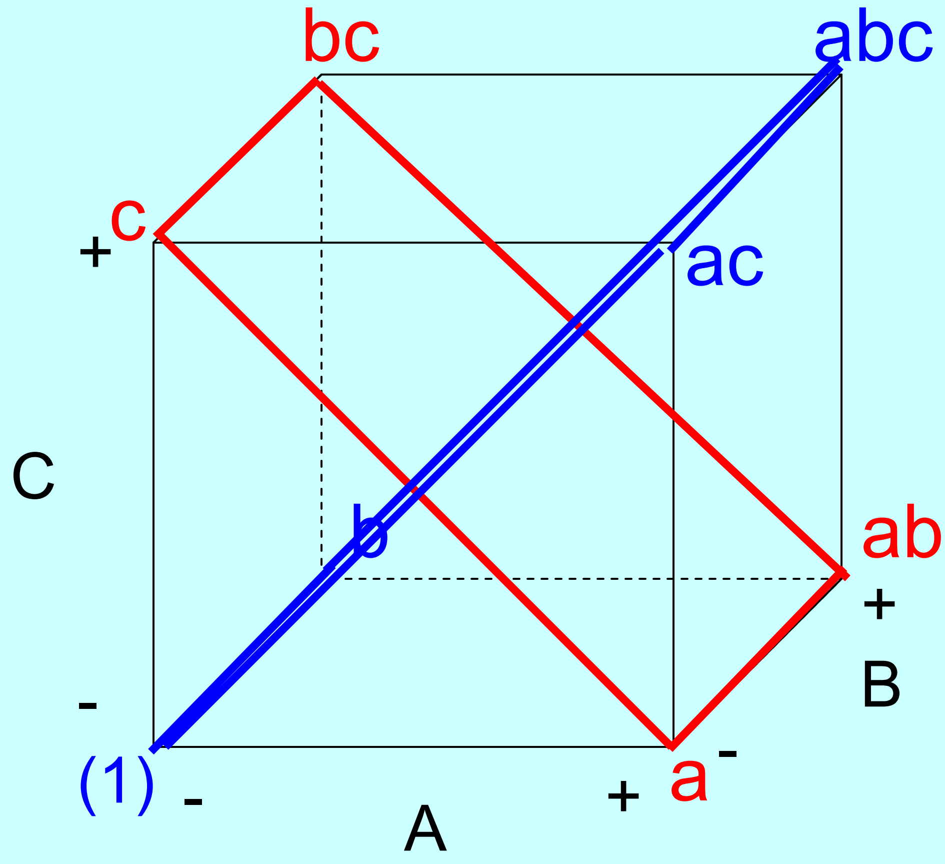
-red plane]



Factorial Efficiency – Graphically (6)

Effect AC

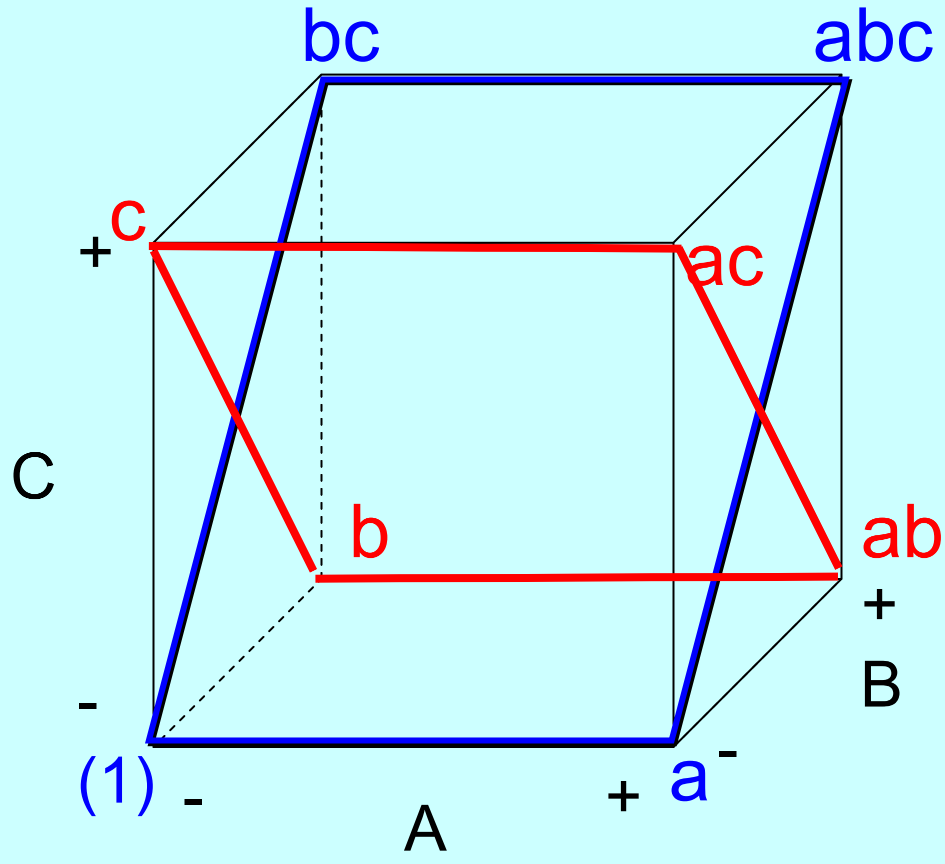
= (1/4*n*) * [blue plane
-red plane]



Factorial Efficiency – Graphically (7)

Effect BC

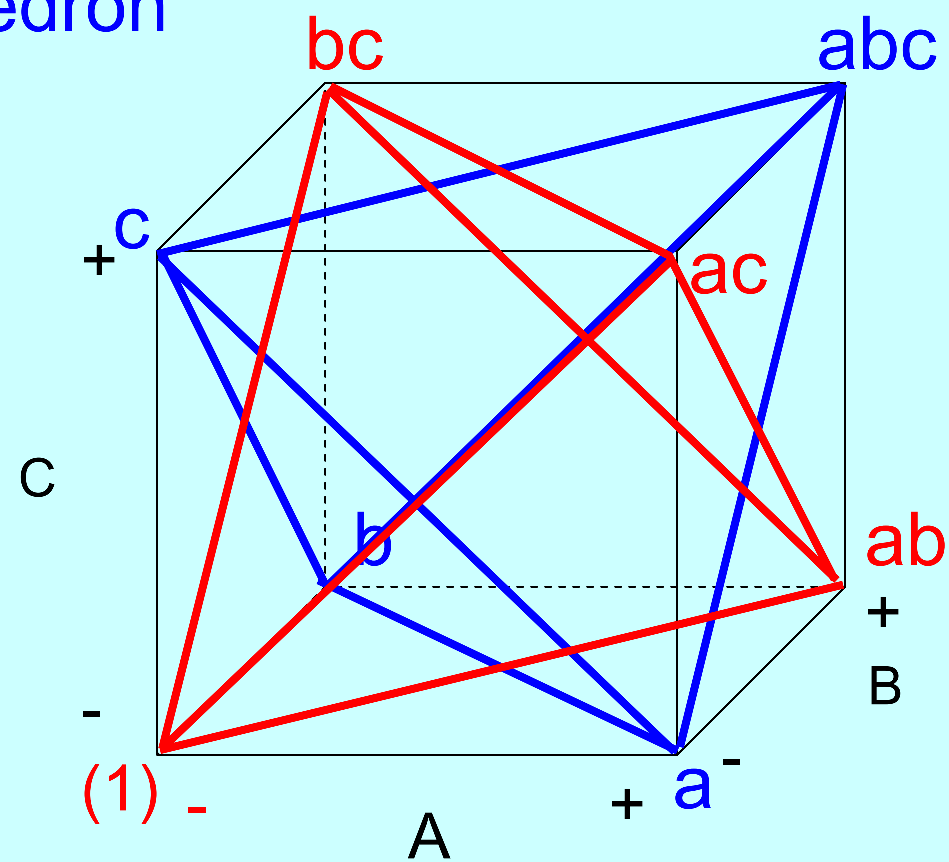
= (1/4*n*) * [blue plane
-red plane]



Factorial Efficiency – Graphically (8)

Effect ABC

= (1/4*n*) * [blue tetrahedron
-red tetrahedron]



Factorial Statistics (1)

- Standard ANOVA table

ANOVA for Selected Factorial Model						
Analysis of variance table [Partial sum of squares]						
Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Model	52202222.22	7	7457460.32	182.57	< 0.0001	significant
A	16885313.28	1	16885313.28	413.38	< 0.0001	
C	30040938.28	1	30040938.28	735.45	< 0.0001	
E	51280.03	1	51280.03	1.26	0.2736	
AC	15268.78	1	15268.78	0.37	0.5467	
AE	2896222.78	1	2896222.78	70.90	< 0.0001	
CE	533286.28	1	533286.28	13.06	0.0014	
ACE	1779912.78	1	1779912.78	43.58	< 0.0001	
Residual	980327.75	24	40846.99			
Cor Total	53182549.97	31				

Factors

Significance Level

Factorial Statistics (2)

DoE Tutorial

- Additional Statistics

Std. Dev.	202.1063818
Mean	8439.46875
C.V.	2.394776115
PRESS	1742804.889

Standard deviation of the main measure

Mean of the main measure

Coefficient of Variation

PRESS = Prediction Error Sum of Squares

- A measure of how well the model will "predict" new data
- Smaller is better but can only be used in a comparative sense

R-Squared	0.98156674
Adj R-Squared	0.97619037
Pred R-Squared	0.96722976
Adeq Precision	42.0620068

Measures the proportion of total variability explained by the model

- R^2 adjusted for the number of Factors
- If non-significant terms are "forced" into the model this can decrease

An estimate of the amount of variability in the new data that would be explained by the full model

- Measures the signal-to-noise ratio in the data
- An indicator if Response Surface Methods (RSM) are applicable
- Values >4 are good

Link to Linear Regression

- Final Model looks like a regression equation

MoM	=
8439.46875	
726.40625	* A
-968.90625	* C
40.03125	* E
-21.84375	* A * C
300.84375	* A * E
-129.09375	* C * E
235.84375	* A * C * E

- Tests of Significance
  - Overall model response
  - Individual coefficients
- Diagnostic tests
  - Residuals
  - Outliers
  - Lack of Fit

Fractional Factorial Designs (1)

- A way to reduce a huge full factorial to something manageable
  - Considerations
    - Required time, resources
    - Complexity of set-up for experiments
  - Major use is in screening experiments where the knowledge of basic effects is not well known
  - If 2^k is very large, may need to run reduced experiment
- Justification
  - Sparsity of Effects – in general, even complex systems are usually driven by a few main effects and low-level interactions
  - Projection Property – fractional factorial designs can be "projected" into larger designs in the subset of significant factors
  - Sequential Experimentation – can combine runs of 2 or more fractional designs into larger designs

Fractional Factorial Designs (2)

- Issue:
  - Confounding of Effects (also called "aliasing") $\Rightarrow$ reduced experiments do *not* evaluate all levels of the factors and their interactions
  - Some mixture of effects is "confounded" and not identifiable
- Challenge:
  - To select the best combination of test elements that stands a reasonable chance of revealing the true effects
  - Alias the (most likely) insignificant or unwanted factors
- Symbology
  - 2^{k-p} designs

Fractional Factorial Designs (3)

- Resolution $\Rightarrow$ a measure of confounding
  - Resolution III
    - No main effect aliased with any other main effect
    - Main effects are aliased with 2-factor interactions
    - 2-factor interactions may be aliased with each other

$$2_{III}^{k-p}$$

  - Resolution IV
    - No main effect aliased with any other main effect
    - No main effect aliased with 2-factor interactions
    - 2-factor interactions may be aliased with each other

$$2_{IV}^{k-p}$$

  - Resolution V
    - No main effect aliased with any other main effect
    - No main effect aliased with 2-factor interactions
    - No 2-factor interactions may be aliased with each other
    - 2-factor interactions are aliased with 3-factor interactions

$$2_{V}^{k-p}$$

Lockheed Martin Center for Innovation

Resolution Trade-offs

MORS

Number of Factors

Experiments	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	Full	1/2 Fract.												
8		Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.								
16			Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.	1/32 Fract.	1/64 Fract.	1/128 Fract.	1/256 Fract.	1/512 Fract.	1/1024 Fract.	1/2048 Fract.
32				Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.	1/32 Fract.	1/64 Fract.	1/128 Fract.	1/256 Fract.	1/512 Fract.	1/1024 Fract.
64					Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.	1/32 Fract.	1/64 Fract.	1/128 Fract.	1/256 Fract.	1/512 Fract.
128						Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.	1/32 Fract.	1/64 Fract.	1/128 Fract.	1/256 Fract.
256							Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.	1/32 Fract.	1/64 Fract.	1/128 Fract.

- Green = Resolution V
- Yellow = Resolution IV
- Red = Resolution III

Lockheed Martin Center for Innovation

Half Replicate/Folding

DoE Tutorial

STD	RUN	A	B	C	D	E
1	7	-1	-1	-1	-1	1
2	4	1	-1	-1	-1	-1
3	6	-1	1	-1	-1	-1
4	10	1	1	-1	-1	1
5	3	-1	-1	1	-1	-1
6	14	1	-1	1	-1	1
7	5	-1	1	1	-1	1
8	1	1	1	1	-1	-1
9	8	-1	-1	-1	1	-1
10	2	1	-1	-1	1	1
11	11	-1	1	-1	1	1
12	15	1	1	-1	1	-1
13	9	-1	-1	1	1	1
14	13	1	-1	1	1	-1
15	12	-1	1	1	1	-1
16	16	1	1	1	1	1

Other Design Variations (1)

- Three levels for k-factors (3^k) designs
- Fractional 3-level designs (3^{k-p})
- Adding Center runs to
  - Get estimates of process variability
  - Gain familiarity with the process
  - Identify system performance limits
- Mixture Designs – where one or more factors are constrained to add to something
  - Usually have constraints like: $x_1 + x_2 + x_3 + \ldots + x_p = 1$
  - Example: A mixture of contributing probabilities
- Nested and Split-Plot designs for experiments with random factors

Other Design Variations (2)

- Irregular Fraction
  - Usually a Resolution V design for 4 to 9 factors where each factor is varied over only 2 levels
  - Two-factor interactions aliased with three-factor and higher
  - Excellent to reduce number of runs and still get clean results
- General Factorial
  - For 1 to 12 factors where each factor may have a different number of levels
  - All factors treated as categorical
- D-Optimal
  - A special design for categorical factors based on a analyst-specified model
  - Design will be a subset of the possible combinations
  - Generated to minimize the error associated with the model coefficients

Other Design Variations (2)

- Plackett-Burman
  - Specialized design for 2 to 31 factors where each factor is varied over only 2 levels
  - Use only if you can reasonably assume NO two-factor interactions; otherwise, use fractional factorial designs
- Taguchi OA
  - Saturated orthogonal arrays – all main effects and NO interactions
  - Special attention must be paid to the alias structure for proper interpretation at both the design phase (prior to runs) and during final analysis

Design of Experiments

Practical Examples

Steps in DoE

- Design the experiment
- Evaluate the design
  - Model specification
  - Power calculations ($1-\beta$)
  - Graphical examination of the standard error of the design
- Conduct the experiment and collect data
- Analyze the results
  - Examine data for transformation suggestions
  - Compute the effects
  - Perform ANOVA
  - Critical!!! – ALWAYS check the diagnostics
  - Examine graphical findings
  - Finalize the analysis

Diagnostics

- Diagnostic steps are the most often omitted – to the analyst's potential embarrassment
- Which of these ANOVA tables are to be believed?

A

Term	DF	Sum of Squares	Mean Square	F Value	Prob > F	% Contribution
Intercept						
A	4	75.69	18.92	2.75	0.0647	40.76
Lack Of Fit	16	110.03	6.88			59.24
Pure Error	0	0.000				0.000
Residuals	16	110.03	6.88			

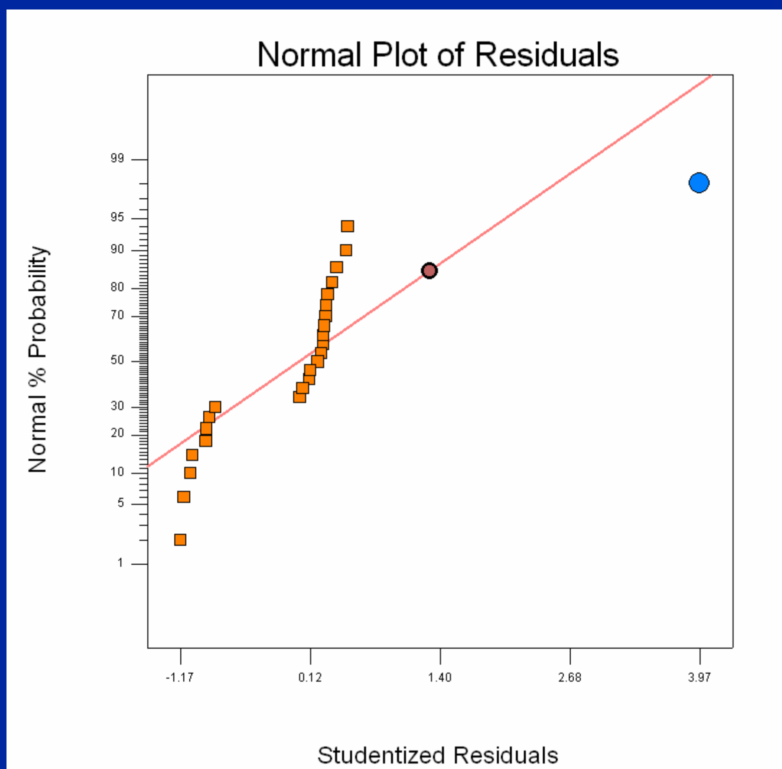
B

Term	DF	Sum of Squares	Mean Square	F Value	Prob > F	% Contribution
Intercept						
A	4	11.68	2.92	25.73	< 0.0001	87.28
Lack Of Fit	15	1.70	0.11			12.72
Pure Error	0	0.000				0.000
Residuals	15	1.70	0.11			

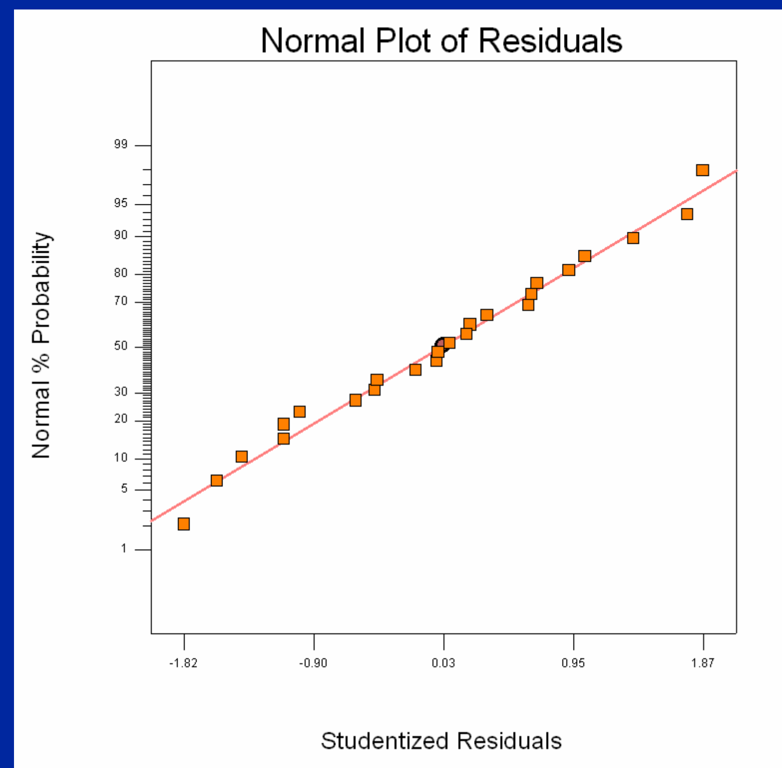
Normality Check

DoE Tutorial

A



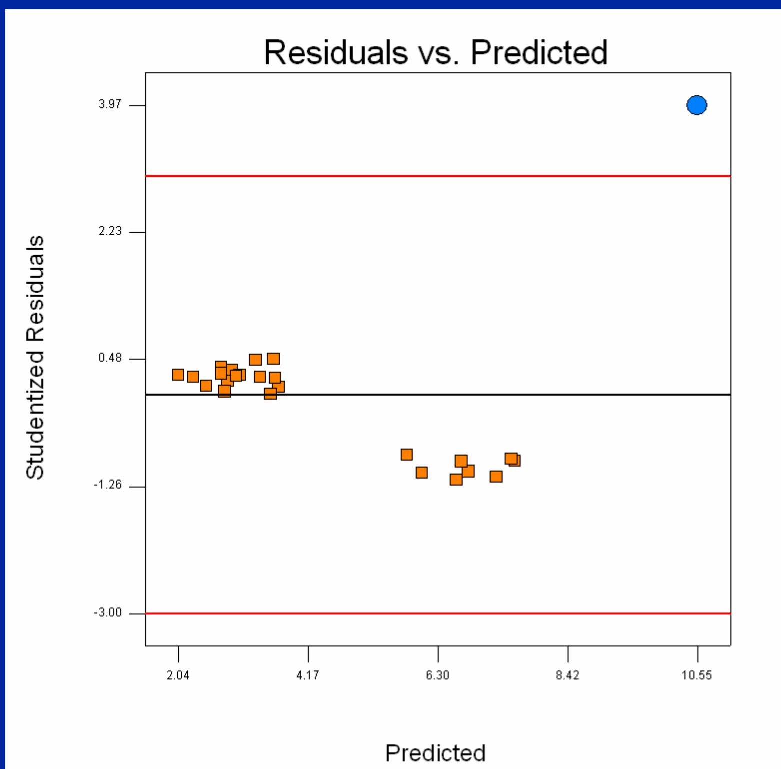
B



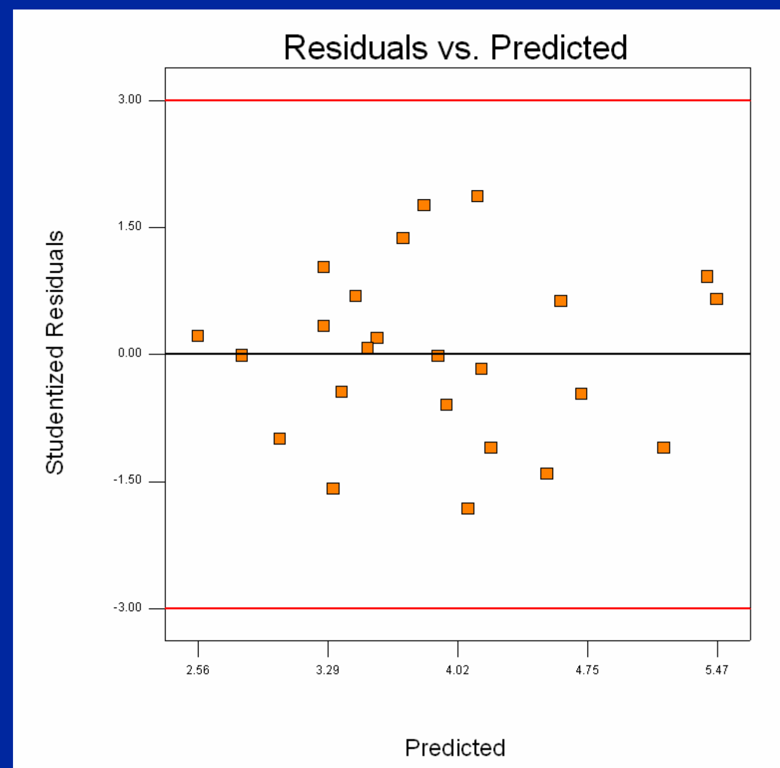
Residuals Check

DoE Tutorial

A



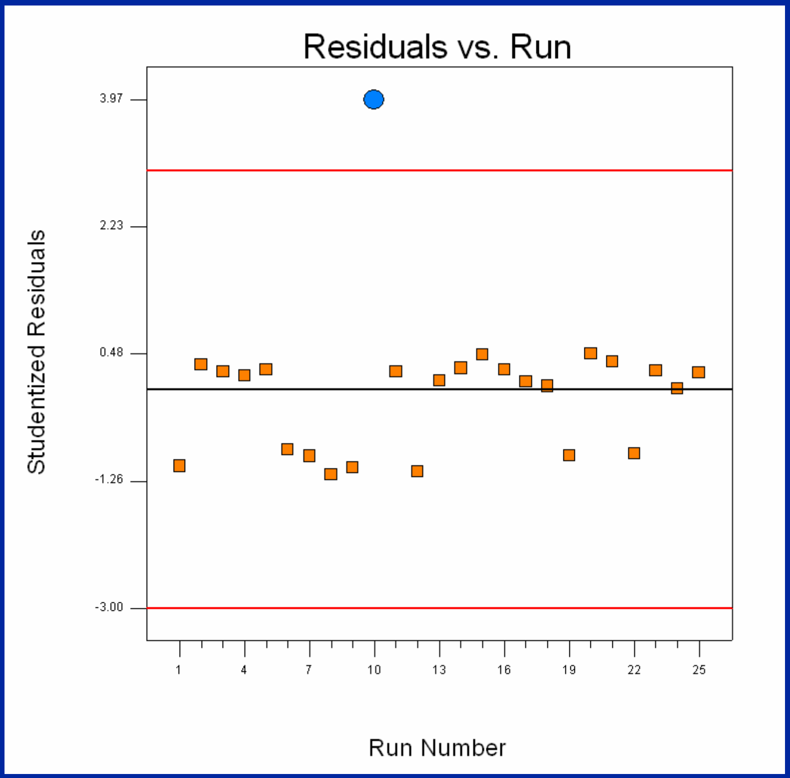
B



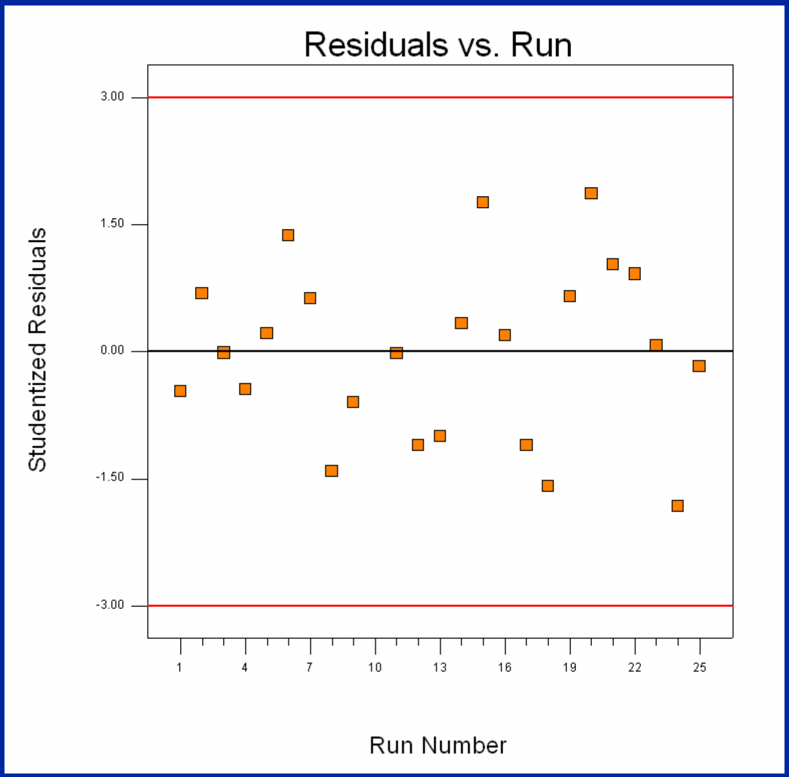
Residuals vs. Run

DoE Tutorial

A



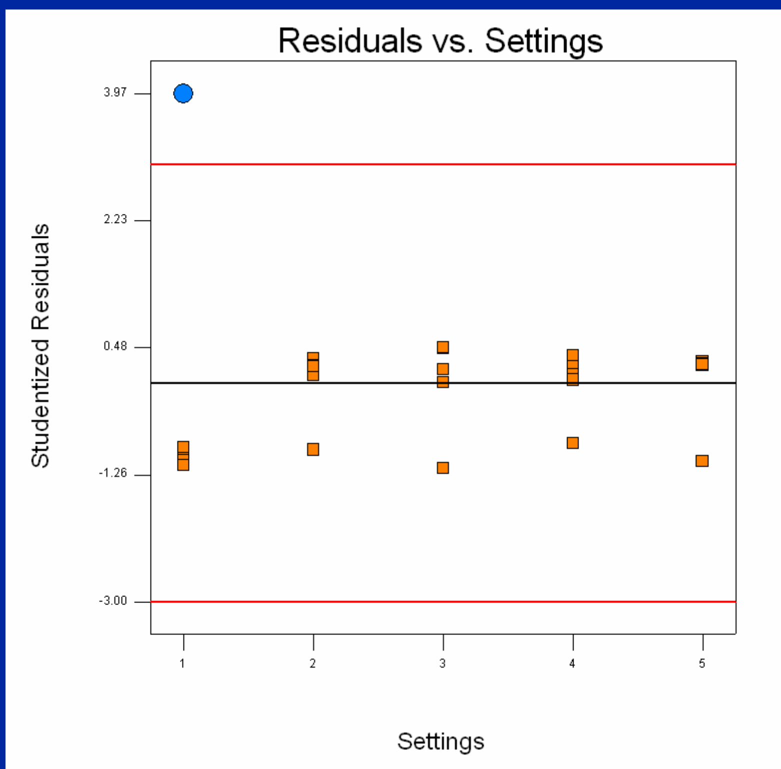
B



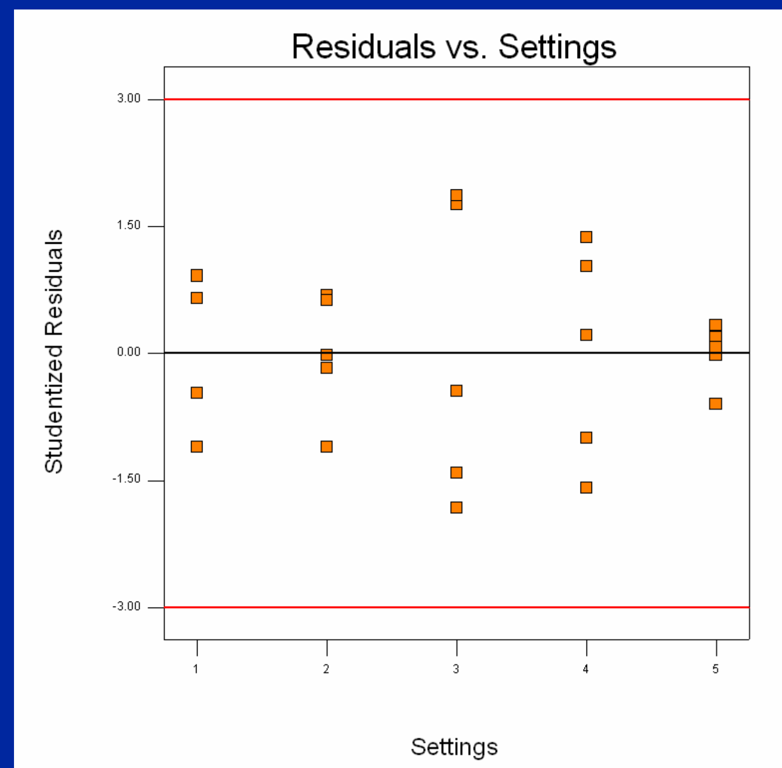
Residuals vs. Settings

DoE Tutorial

A



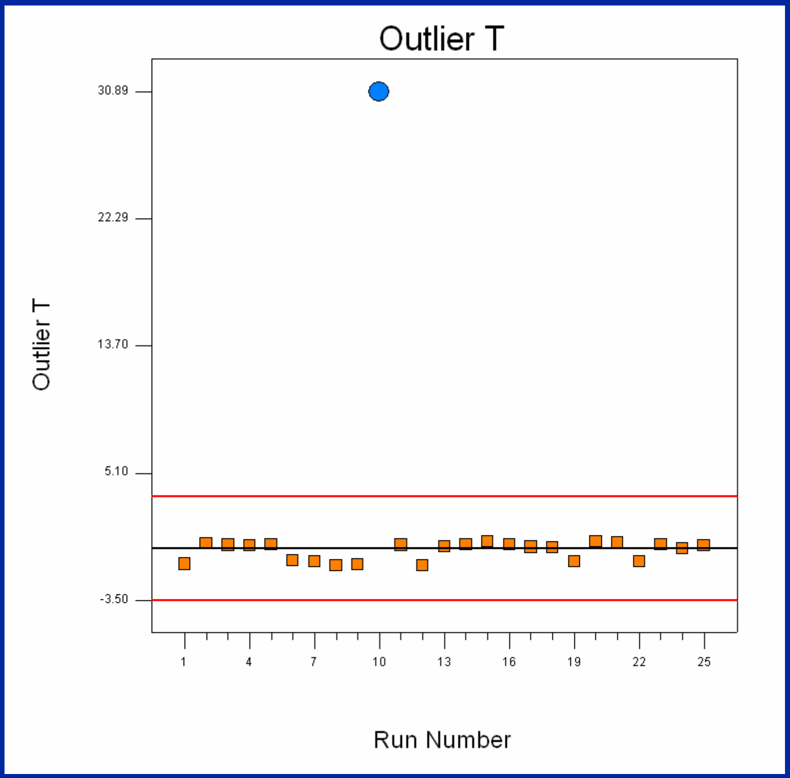
B



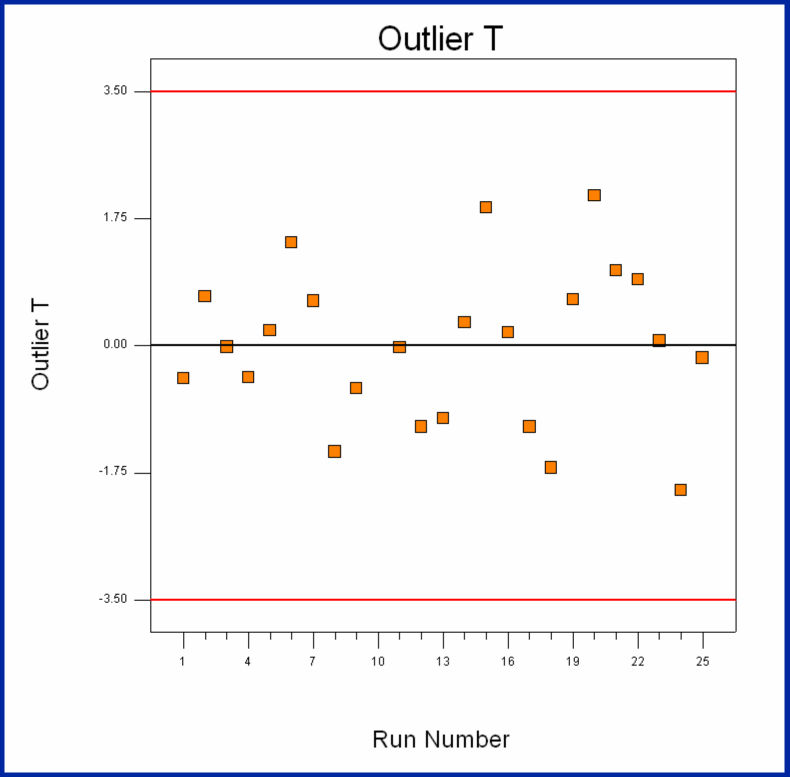
Outlier T Check

DoE Tutorial

A

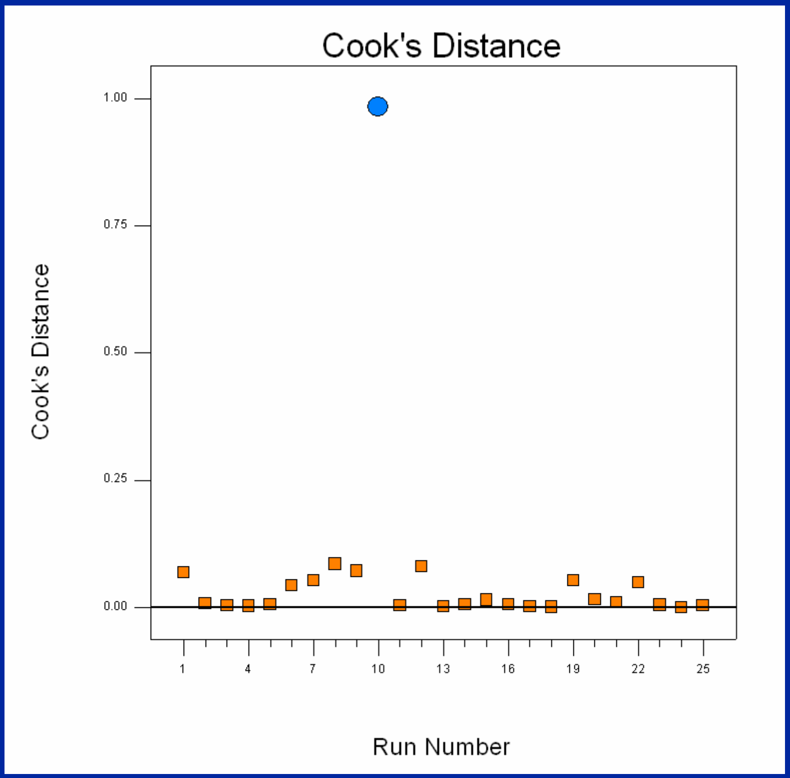


B

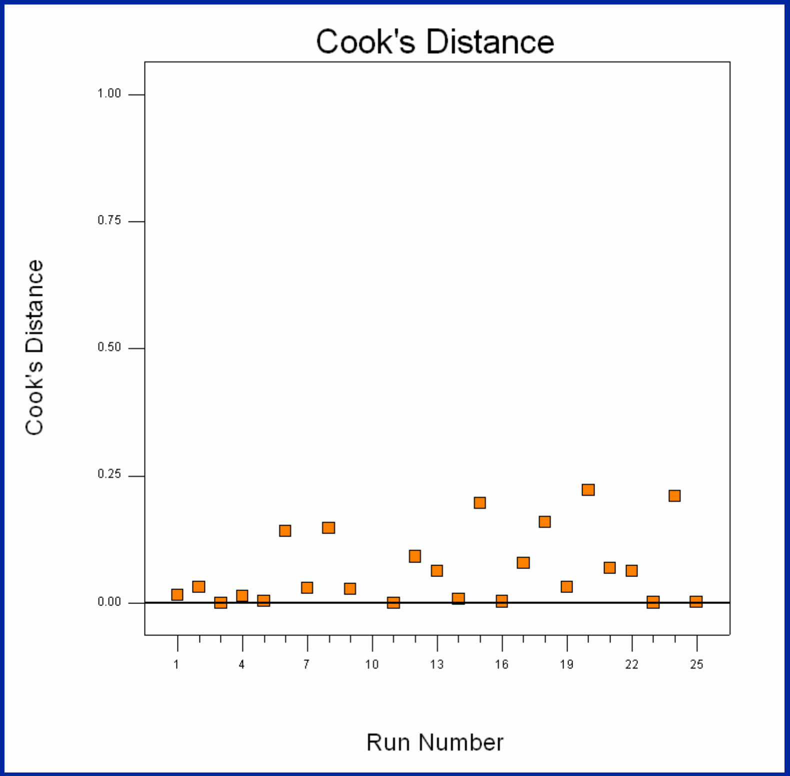


Cook’s Distance

A



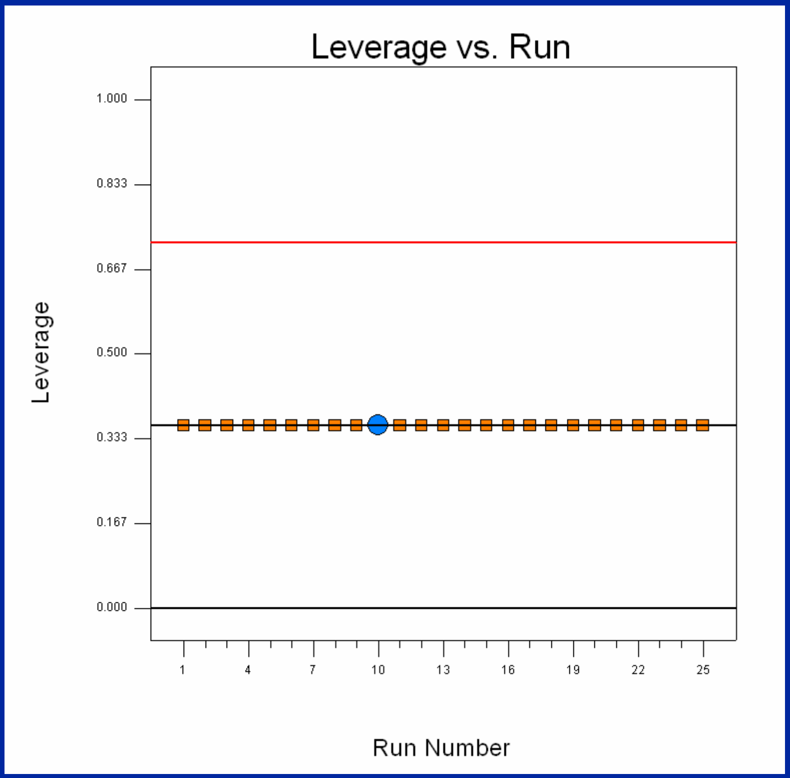
B



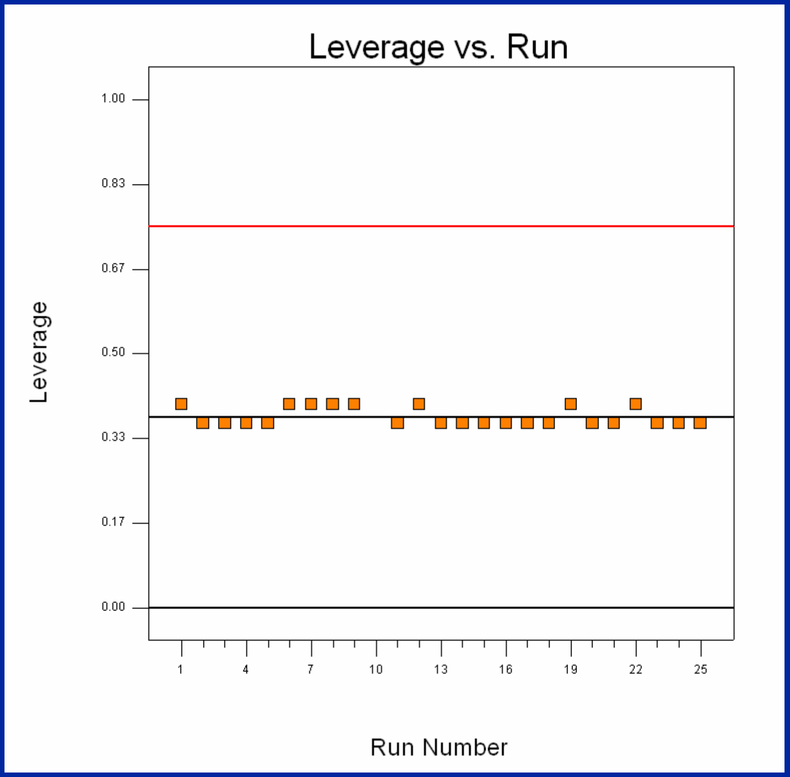
Leverage

DoE Tutorial

A



B

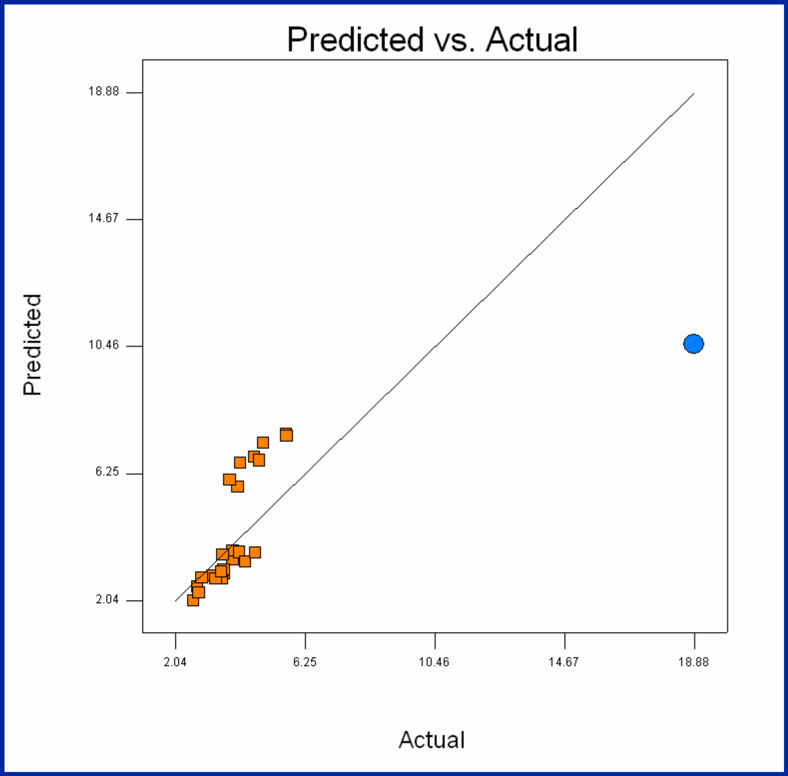


Lockheed Martin Center for Innovation

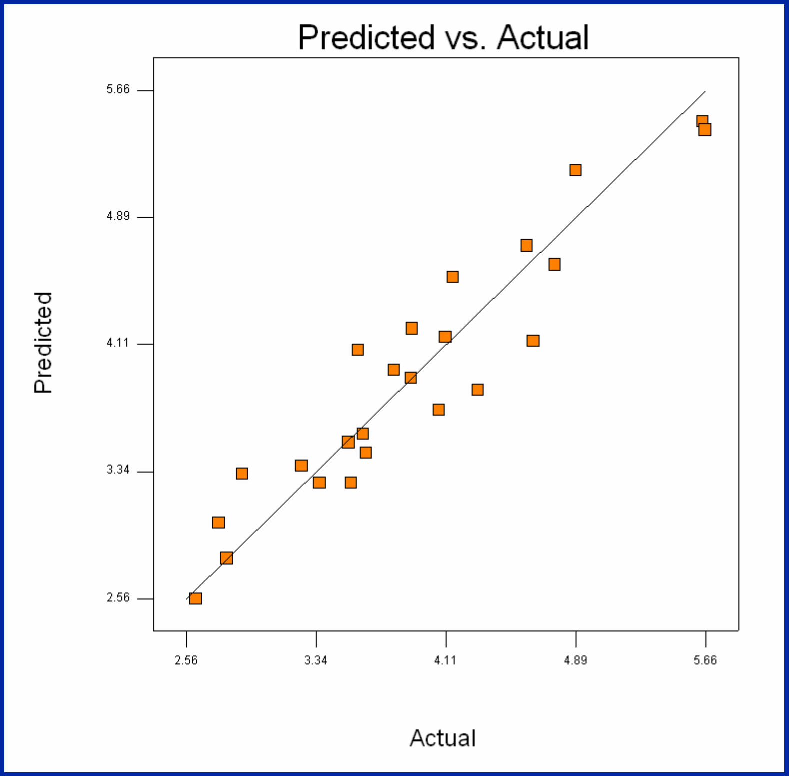
Predicted vs. Actual

DoE Tutorial

A



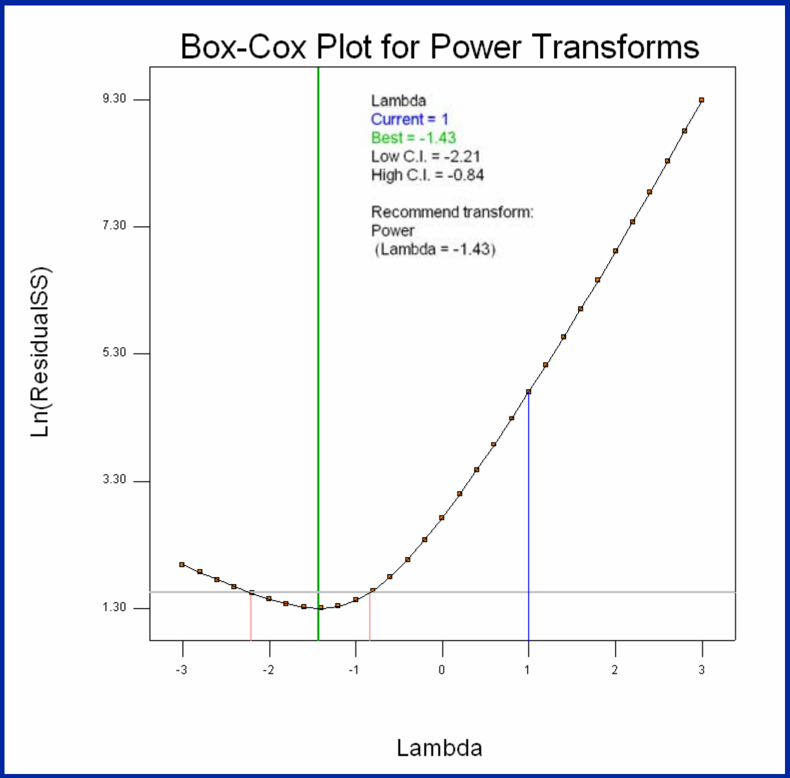
B



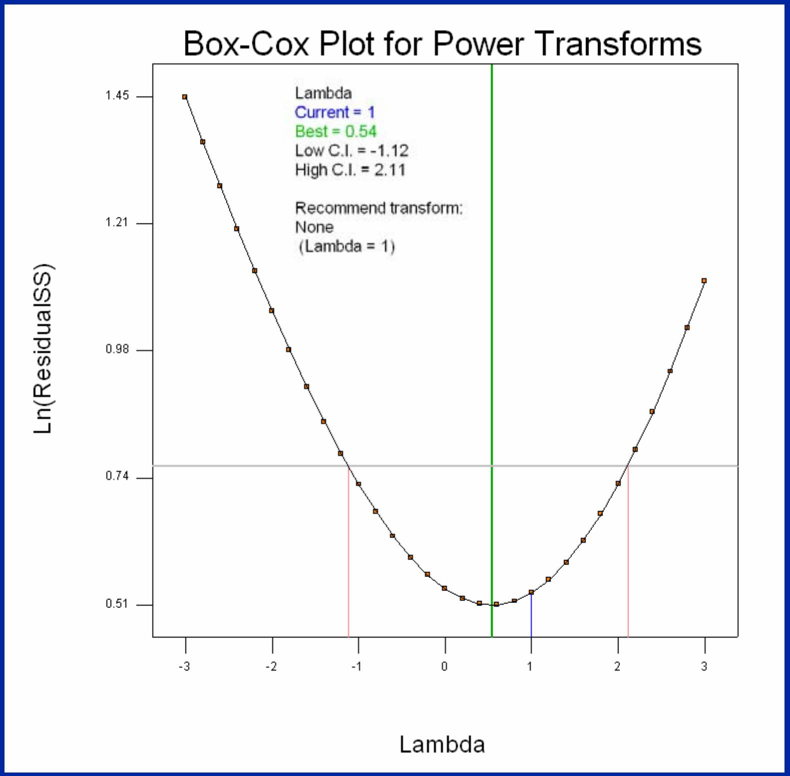
Power Transform Recommendation

DoE Tutorial

A



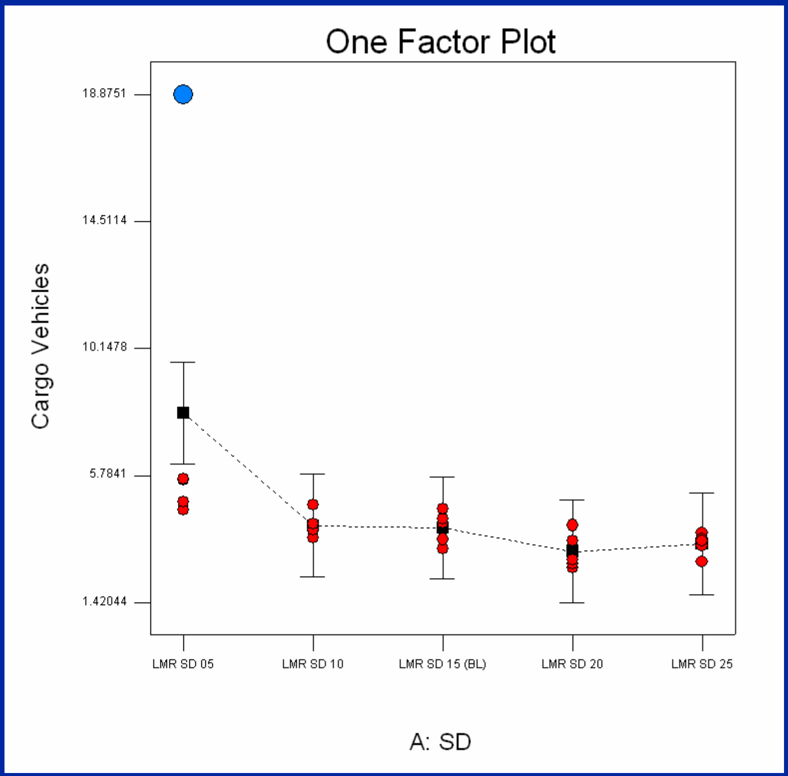
B



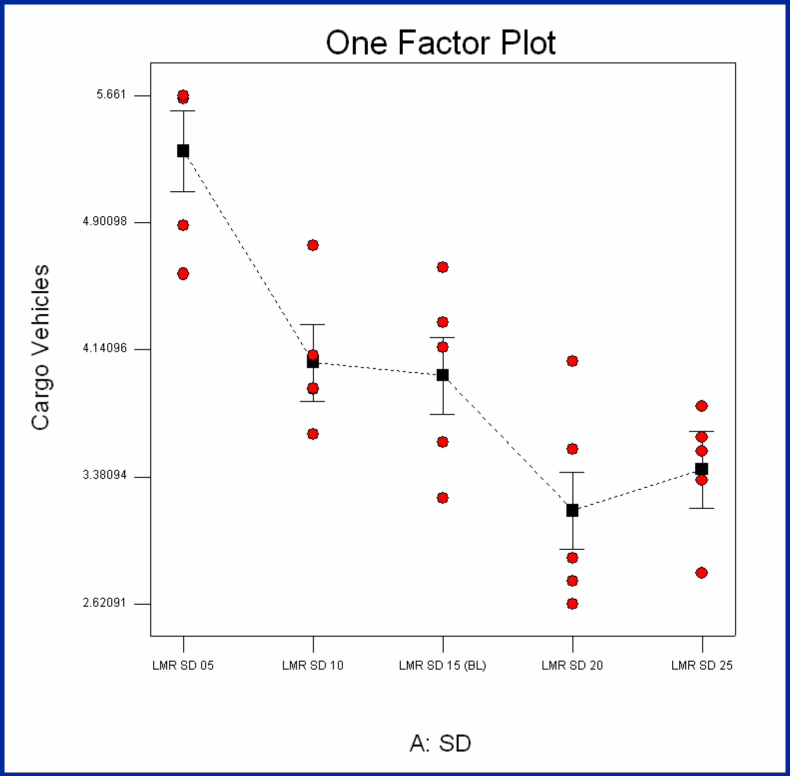
Single Factor Plot

DoE Tutorial

A



B



Remember!!

ALWAYS perform the diagnostic tests!

Measuring the Effect of C3I on Combat: Methodology and Results

An Example of the Application of Design of Experiments Concepts and Techniques

Initial Guidance

- Evaluate the impact of the representations of C3I on combat outcomes in a campaign-level force-on-force model.
  - Perform sensitivity analyses across all areas of C3I, utilizing the existing test scenario.
  - Determine which C3I-related input data have the most impact on combat outcomes.

Approach

- Select specific C3I functions to be examined.
- Design the Experiment.
- Prepare model software and scenario.
- Execute model runs.
- Analyze the output and report findings.

Excluded Elements

The following items are not part of the study as they are either not controllable from a military sense or they represent different tactics or behaviors, which are not of interest for this study:

- Weather
- Intelligence Ratings
- Force Structure
- ISR Collection Plans

Experimental Design

- Bundled multiple factors into 3 categories to describe C3I functionality in terms of:
  - Timeliness
  - Quantity
  - Quality

Statistical Considerations

- Each candidate factor could influence combat outcome either:
  - By itself,
  - In concert with another factor,
  - In opposition to another factor.
- Previous research in this area has shown serious non-linear effects.

Formal Design

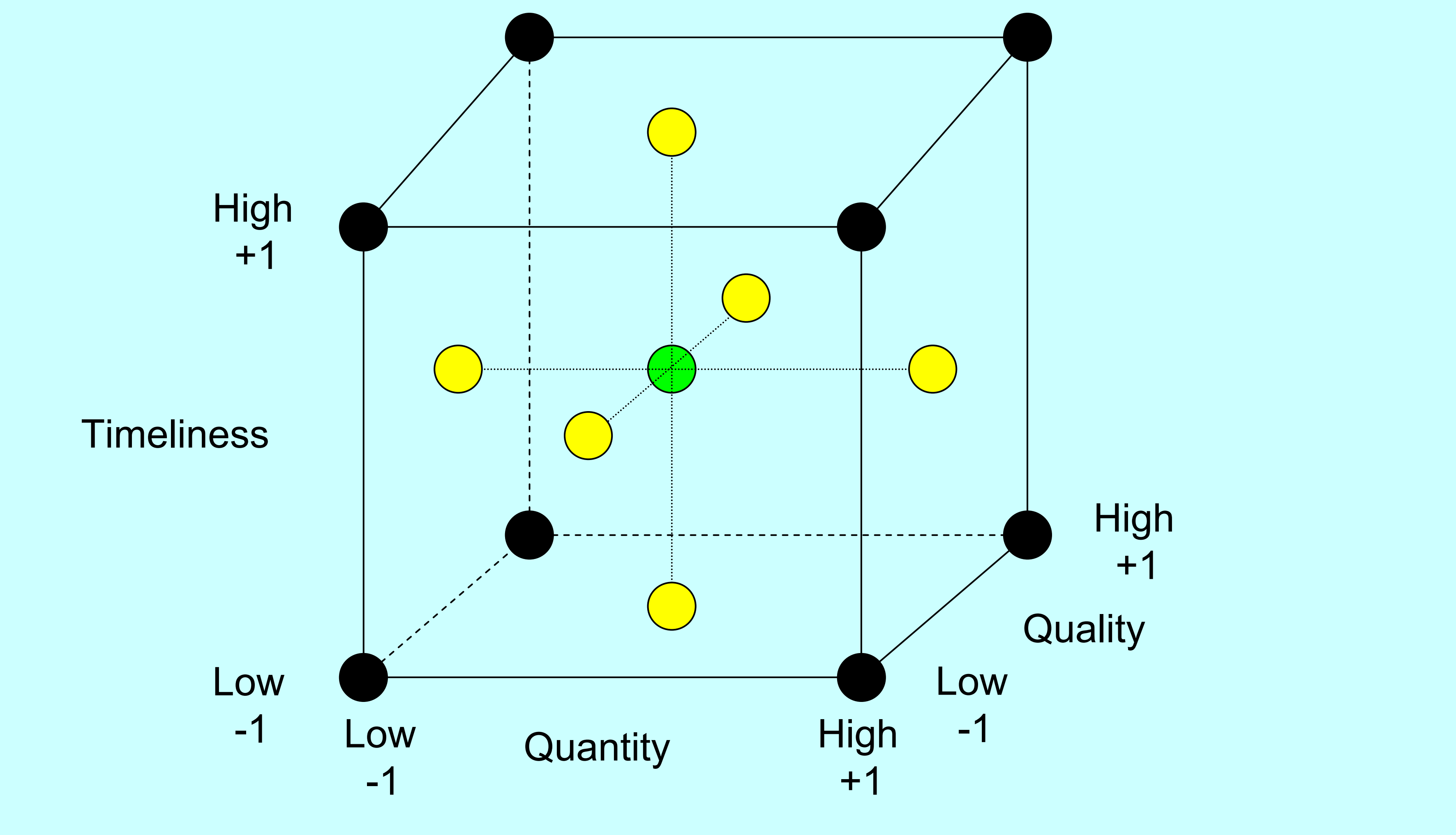
- Three-factor design with interaction and non-linear terms.

$$CO = \left(\begin{array}{c} \boxed{\beta_0} + \boxed{\beta_1 T + \beta_2 Q_T + \beta_3 Q_L} + \\ + \boxed{\beta_{11} T^2 + \beta_{22} Q_T^2 + \beta_{33} Q_L^2} \\ + \boxed{\beta_{12} T Q_T + \beta_{13} T Q_L + \beta_{23} Q_T Q_L + \beta_{123} T Q_T Q_L} + \boxed{\varepsilon} \end{array} \right)$$

where:

- CO = Combat Outcome
- T = Timeliness
- Q_T = Quantity
- Q_L = Quality
- β_i = an unknown value to be estimated
- ε = the error term

Measurement Points



Face-Centered Central Composite Design

Design Matrix

- The FC-CCD yields the following design matrix:

$$CO = \left[\underbrace{\begin{matrix} -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 \end{matrix}}_{\text{Corner Points}} \; \underbrace{\begin{matrix} 0 & 0 & 0 & 0 & -1 & +1 \\ 0 & 0 & -1 & +1 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 & 0 \end{matrix}}_{\text{Face Points}} \; \underbrace{\begin{matrix} 0 \\ 0 \\ 0 \end{matrix}}_{\text{CENTER}}\right]^T$$

Parameter Settings

- Each parameter was chosen so that 3 settings were possible to match the FC-CCD requirements:
  - High (meaning improved or enhanced performance)
  - Center (baseline)
  - Low (meaning reduced or degraded performance)

Timeliness (T) Settings

Parameters	Low (-1)	Center (0)	High (+1)
Reporter Delay Time (RDT)	8	4	0
Presented Communications Load (PCL)	$1.25*PCL_{Base}$	PCL_{Base}	$0.75*PCL_{Base}$
Maximum Communications Network Capacity (MCNC)	$0.75*MCNC_{Base}$	$MCNC_{Base}$	$1.25*MCNC_{Base}$

Lockheed Martin Center for Innovation

Quantity (Q_T) Settings

Parameters	Low (-1)	Center (0)	High (+1)
IMINT Probability of Detection ($P_{d\text{-}IMINT}$)	0.4	0.7	1.0
Sensor Footprint (SFP)	$0.707*SFP_{Base}$	SFP_{Base}	$1.414*SFP_{Base}$
COMINT Sensor Search Rate (CSSR)	$0.5*CSSR_{Base}$	$CSSR_{Base}$	$2*CSSR_{Base}$

Quality (Q_L) Settings

Parameters	Low (-1)	Center (0)	High (+1)
Probability of Correct Classification for MTI sensors ($P_{CC\text{-}MTI}$)	0.75 : 0.25	0.5 : 0.5	0.25 : 0.75
Quality Probability for explicit IMINT search ($P_{Q\text{-}IMINT}$)	$P_{Degrade\text{-}Q\text{-}IMINT}$	$P_{Q\text{-}IMINT}$	$P_{Upgrade\text{-}Q\text{-}IMINT}$
Association Threshold (AT)	$0.5*AT_{Base}$	AT_{Base}	$2*AT_{Base}$

Probability Transforms (1)

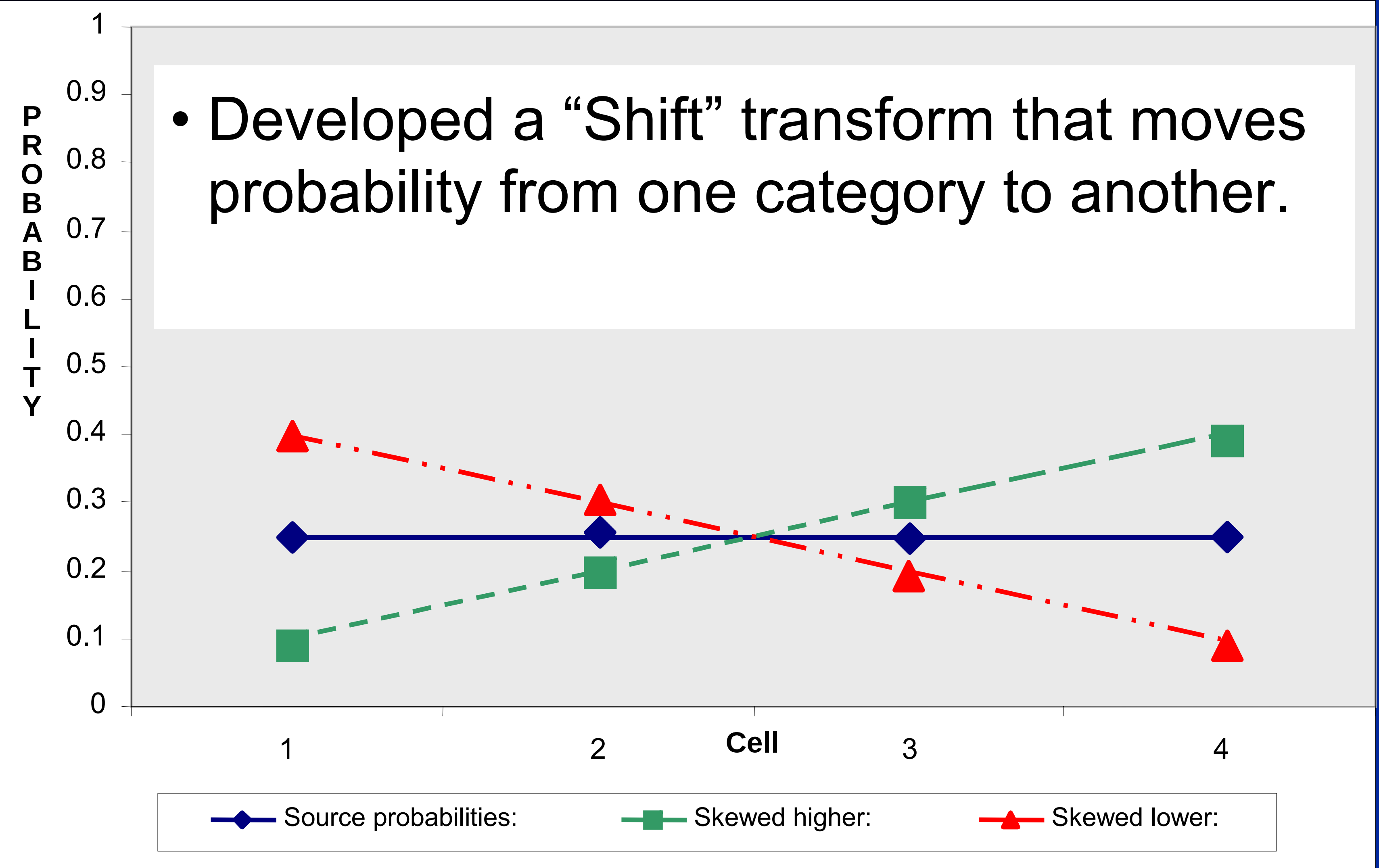
- Quality Classification Probability is actually a distribution, not a single value, where $\Sigma p_i = 1$

Quality Level	1	2	3	N
Probability	p_1	p_2	p_3	p_N

Probability Transforms (2)

DoE Tutorial

- Developed a "Shift" transform that moves probability from one category to another.



Combat Outcome Measures

Combat Outcome	Sources
$R_{K\text{-}All}$	Direct Fire KVSB* Indirect Fire KVSB Air-to-Ground KVSB
$R_{K\text{-}DF}$	Direct Fire KVSB
$R_{K\text{-}IF}$	Indirect Fire KVSB
$R_{K\text{-}A2G}$	Air-to-Ground KVSB

*KVSB = Killer-Victim Scoreboard

Run Results Summary

DoE Tutorial

Response	Name	Observations	Minimum	Maximum
Y1	DF Kills - Red	150	166.696716	543.007375
Y2	IF Kills - Red	150	38.507262	317.413308
Y3	A2G Kills - Red	150	639.270296	1986.671009
Y4	Total Kills - Red	150	1039.098472	2422.615195

DF = Direct Fire

IF = Indirect Fire

A2G = Air-to-Ground

ANOVA – Total Kills

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	$\alpha = 0.05$
Model	6,024,287.18	9	669,365.24	20.2233	< 0.0001	*
T	3,235,042.05	1	3,235,042.05	97.739	< 0.0001	*
Q_T	659,344.82	1	659,344.82	19.9205	< 0.0001	*
Q_L	794,562.33	1	794,562.33	24.0058	< 0.0001	*
T^2	212,422.81	1	212,422.81	6.4178	0.0124	*
$Q_T{}^2$	558,906.79	1	558,906.79	16.886	< 0.0001	*
$Q_L{}^2$	356,637.37	1	356,637.37	10.7749	0.0013	*
TQ_T	17,652.64	1	17,652.64	0.5333	0.4664	
TQ_L	47,410.91	1	47,410.91	1.4324	0.2334	
Q_TQ_L	199,200.94	1	199,200.94	6.0184	0.0154	*
Residual	4,633,827.61	140	33,098.77			
Lack of Fit	660,829.65	5	132,165.93	4.4909	0.0008	*
Pure Error	3,972,997.96	135	29,429.61			
Cor Total	10,658,114.79	149				

Fitted Model – Total Kills

$$CO = \left(\begin{array}{c} 1595.89 + 179.86T + 81.20Q_T + 89.14Q_L + \\ + 90.89T^2 + 147.43Q_T^2 - 117.77Q_L^2 \\ - 14.85TQ_T - 24.34TQ_L - 49.90Q_T Q_L \end{array} \right)$$

Design Evaluation

- Evaluate the formal design by examining the following parameters:
  – Calculate power of the tests
  – Perturbation plots
  – Contour plots
  – Standard error graphs

Power of the Design

DoE Tutorial

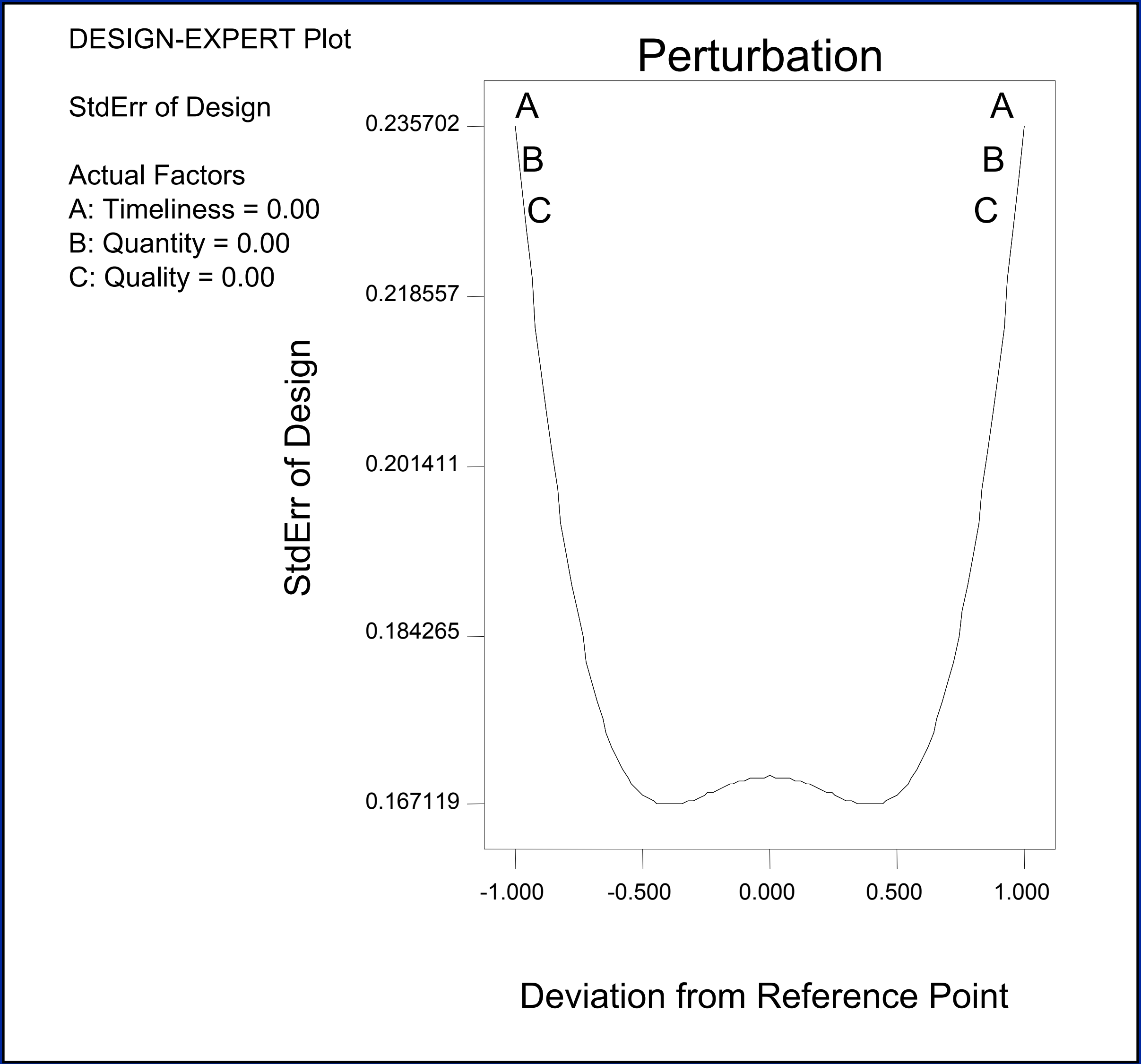
Term	StdErr**	Power at 5% alpha level for effect of: 1/2 Std. Dev.	1 Std. Dev.	2 Std. Dev.
T	0.1	69.90%	99.90%	99.90%
Q_T	0.1	69.90%	99.90%	99.90%
Q_L	0.1	69.90%	99.90%	99.90%
T^2	0.1972027	71.20%	99.90%	99.90%
Q_T^2	0.1972027	71.20%	99.90%	99.90%
Q_L^2	0.1972027	71.20%	99.90%	99.90%
TQ_T	0.1118034	60.30%	99.30%	99.90%
TQ_L	0.1118034	60.30%	99.30%	99.90%
Q_TQ_L	0.1118034	60.30%	99.30%	99.90%

**Basis Std. Dev. = 1.0

Perturbation Plot

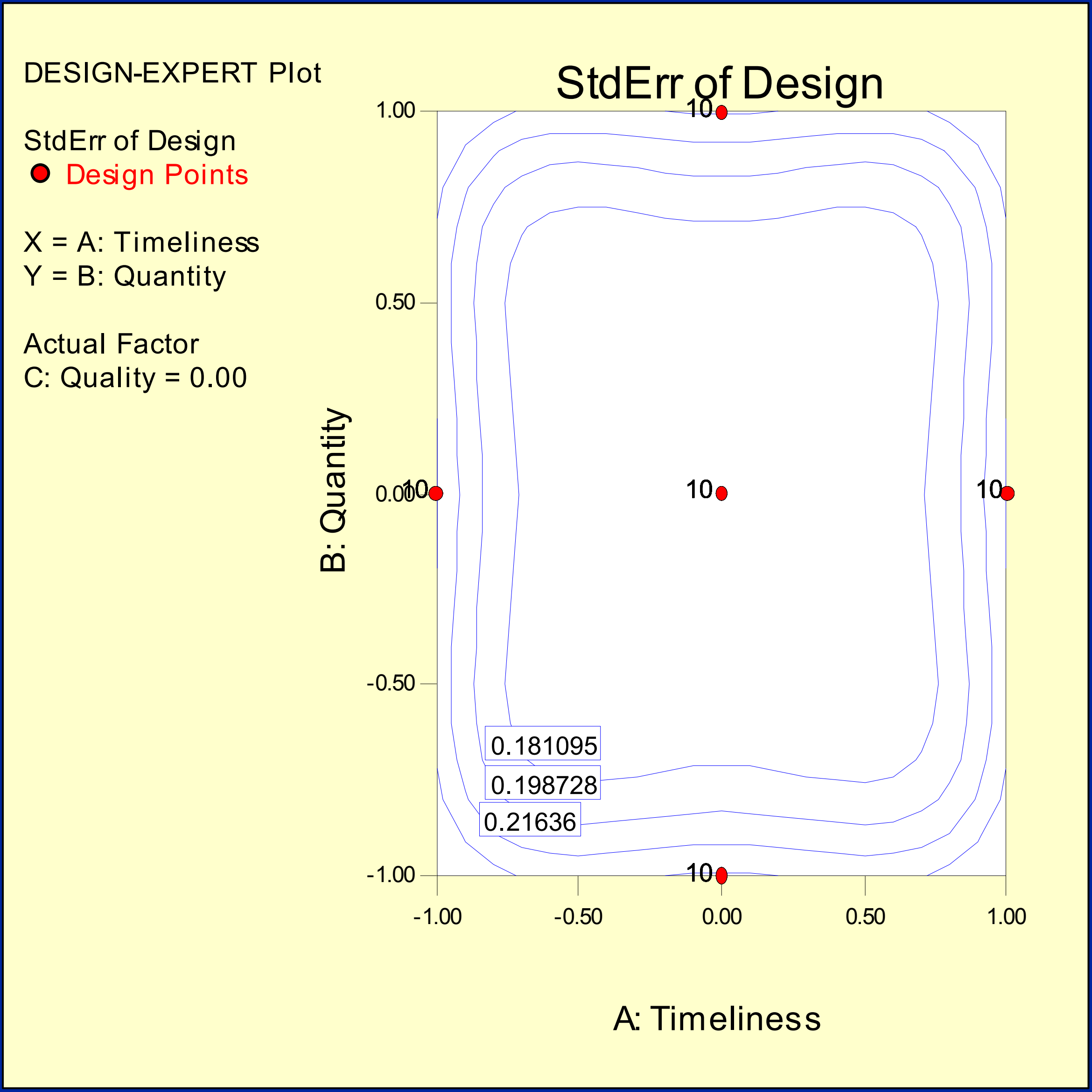
DoE Tutorial

- Plot of Standard Error of Design
  - Shows error of the estimates increases at the edge of the design space
  - All factors overlap: they have the same standard error
  - Conclusions based on extreme values may be subject to major qualification



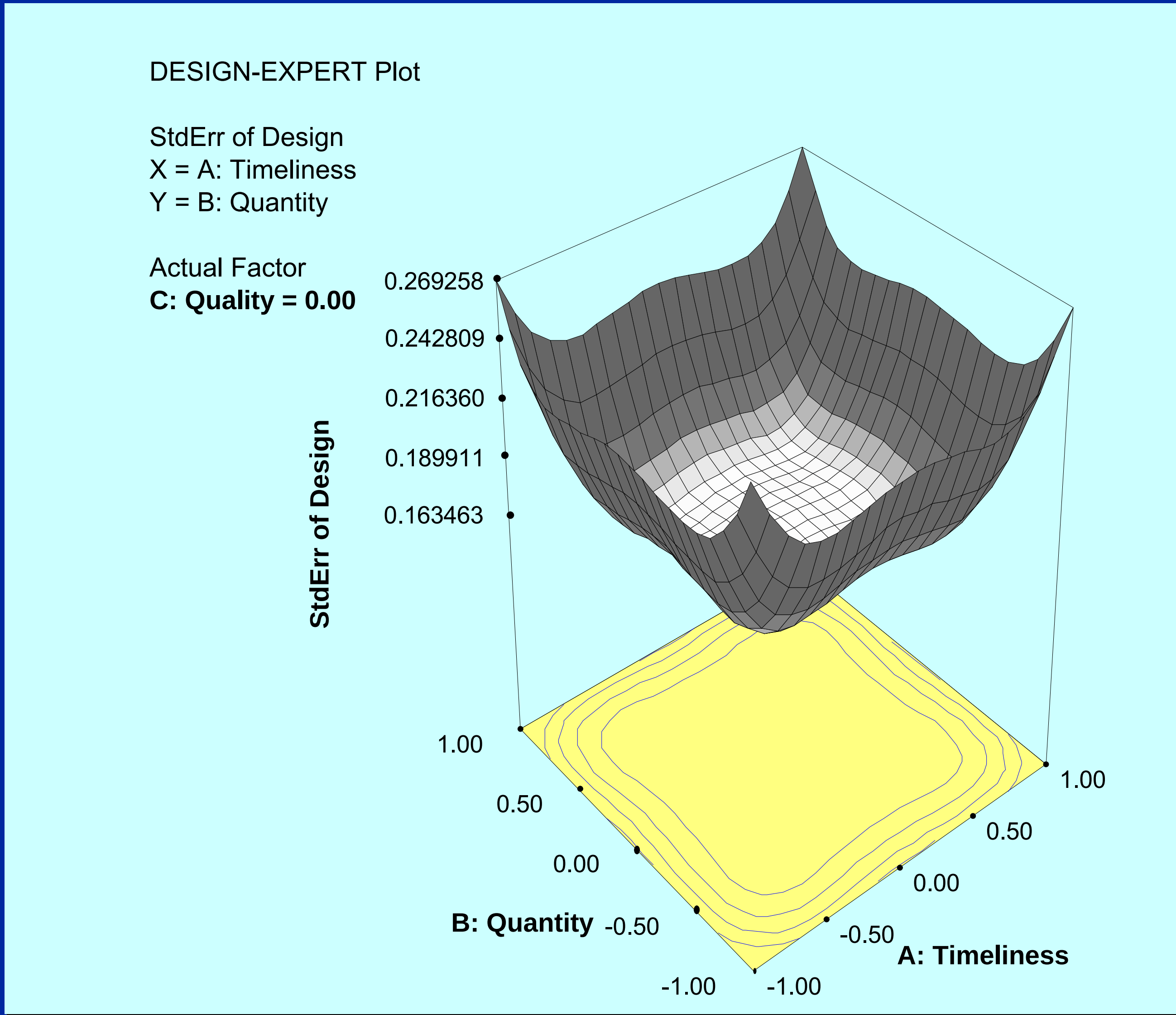
Contour Plot

- Plot of Standard Error of Design
  - 2-Factor view for a constant setting of the 3rd factor
  - Tight contours indicate steepness of response
  - More difficult to read than a 3-D plot



3-D Standard Error Plot

- Plot of Standard Error of Design
  - 3-D, 2-Factor view for a constant setting of the 3rd factor
  - Corresponding contour plot is shown on the base
  - Depth of shading indicates steepness of slope

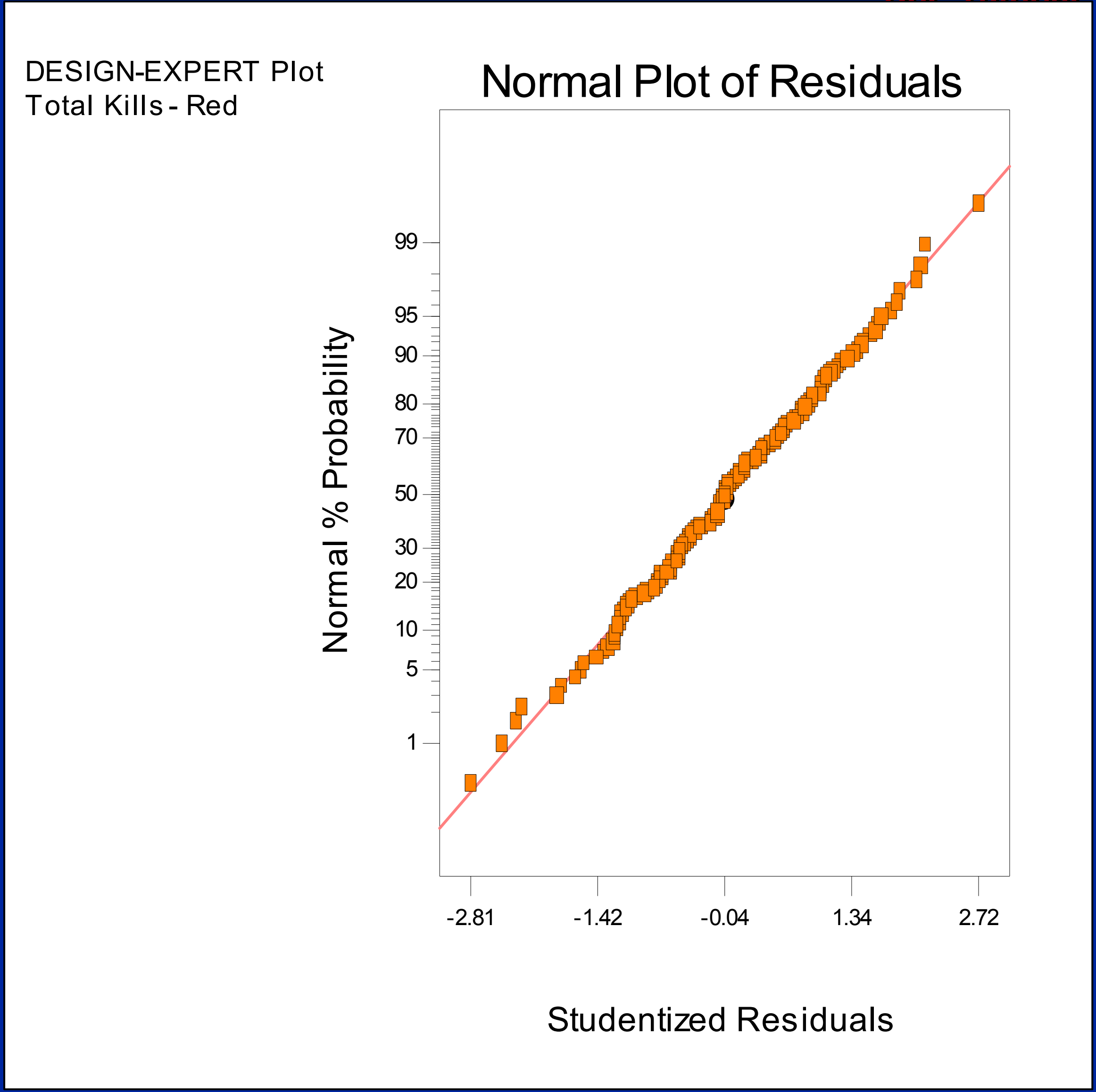


Diagnostic Tests

- Examine data output with:
  - Normal plot of the residuals
  - Residuals vs. predicted error
  - Residuals vs. run
  - Residuals vs. Timeliness
  - Residuals vs. Quantity
  - Residuals vs. Quality
- Conduct outlier investigation
- Conduct transform analysis

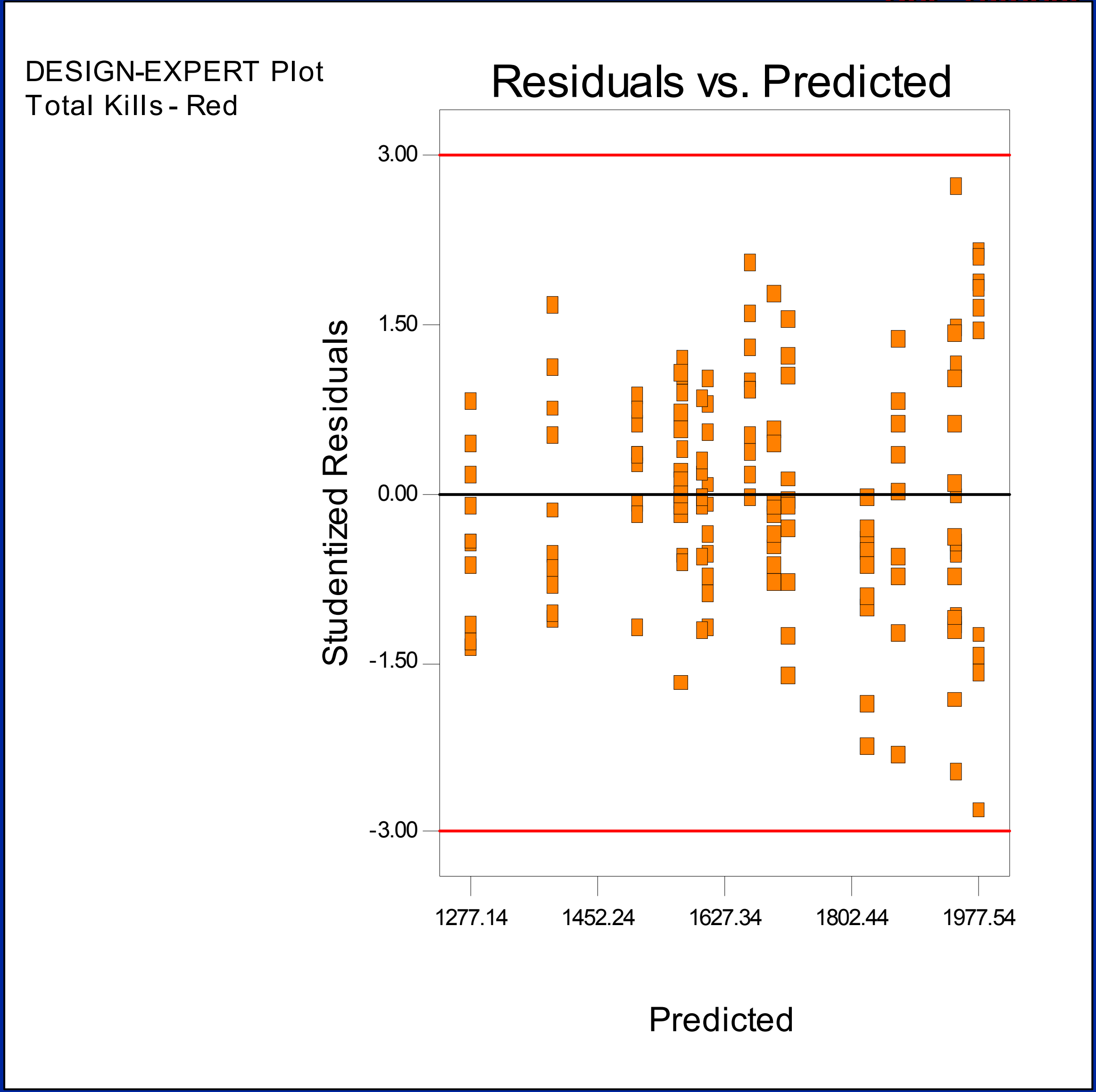
Normal Plot of Residuals

- Residual Plot
  - Desired – data points fall on a straight line
  - Actual – does not show any serious abnormality
  - Results – OK



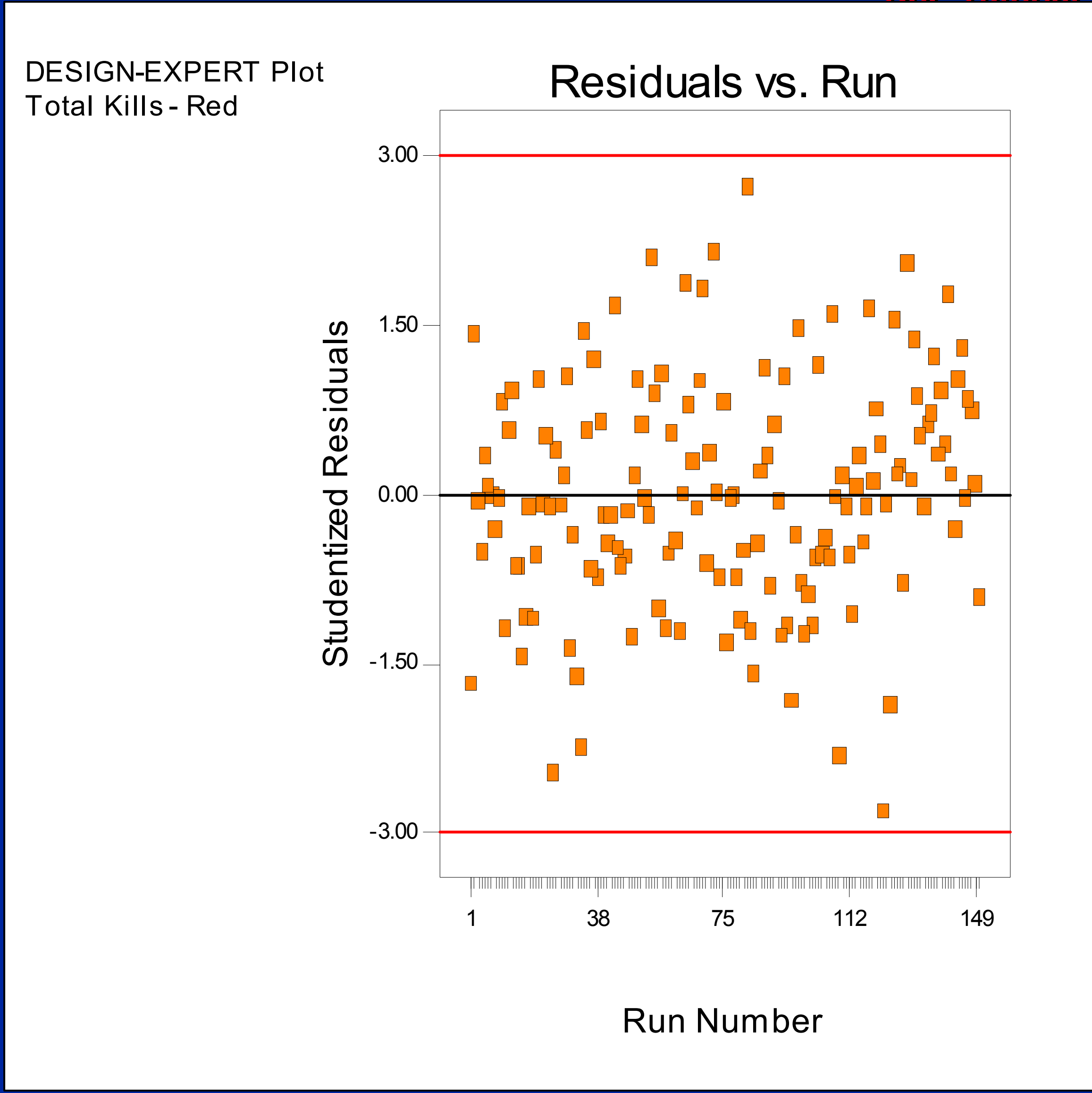
Residuals vs. Predicted

- Residual Analysis
  - Desired – no apparent pattern in the observed data
  - Actual – no pattern in the observed data
  - Results – OK



Residuals vs. Run

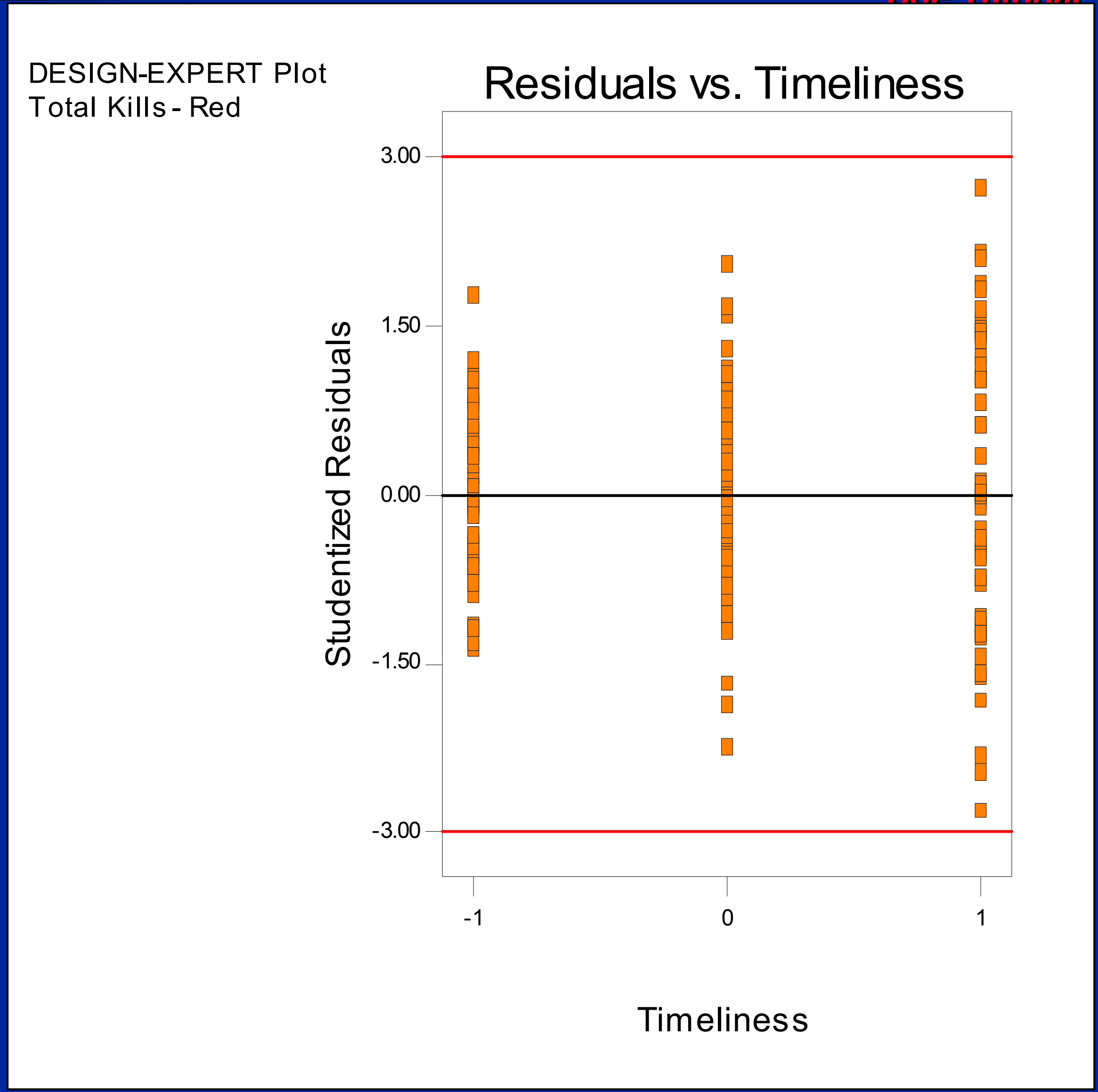
- Residual Analysis
  - Desired – no apparent pattern in the observed data
  - Actual – no pattern in the observed data
  - Results – OK



Residuals vs. Timeliness

DoE Tutorial

- Residual Analysis
  - Desired – no apparent pattern in the observed data
  - Actual – no pattern in the observed data
  - A slight expansion as settings shift from Low to High but not strong enough to invalidate results
  - Results – OK



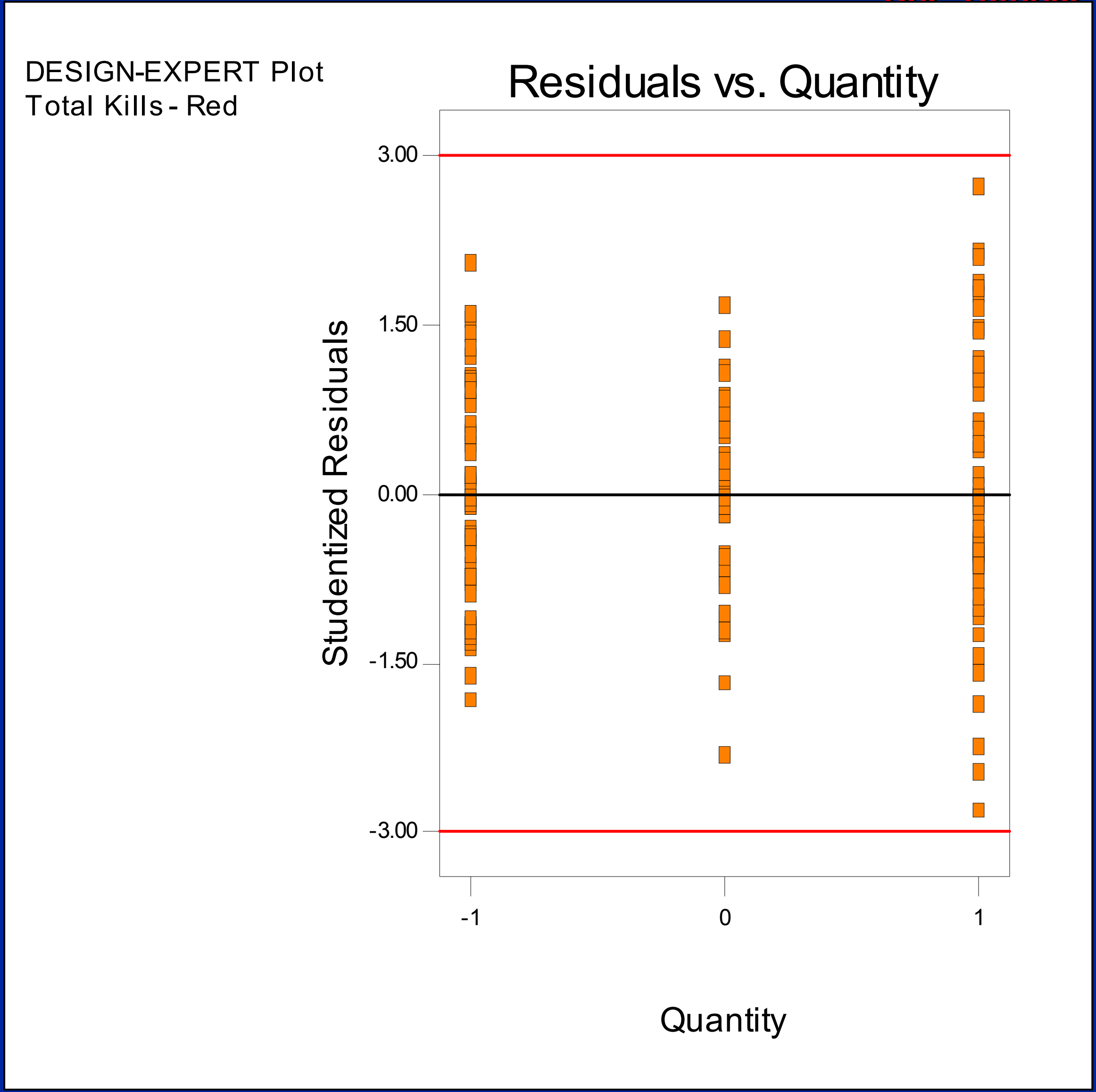
Lockheed Martin Center for Innovation

Residuals vs. Quantity

MORS

DoE Tutorial

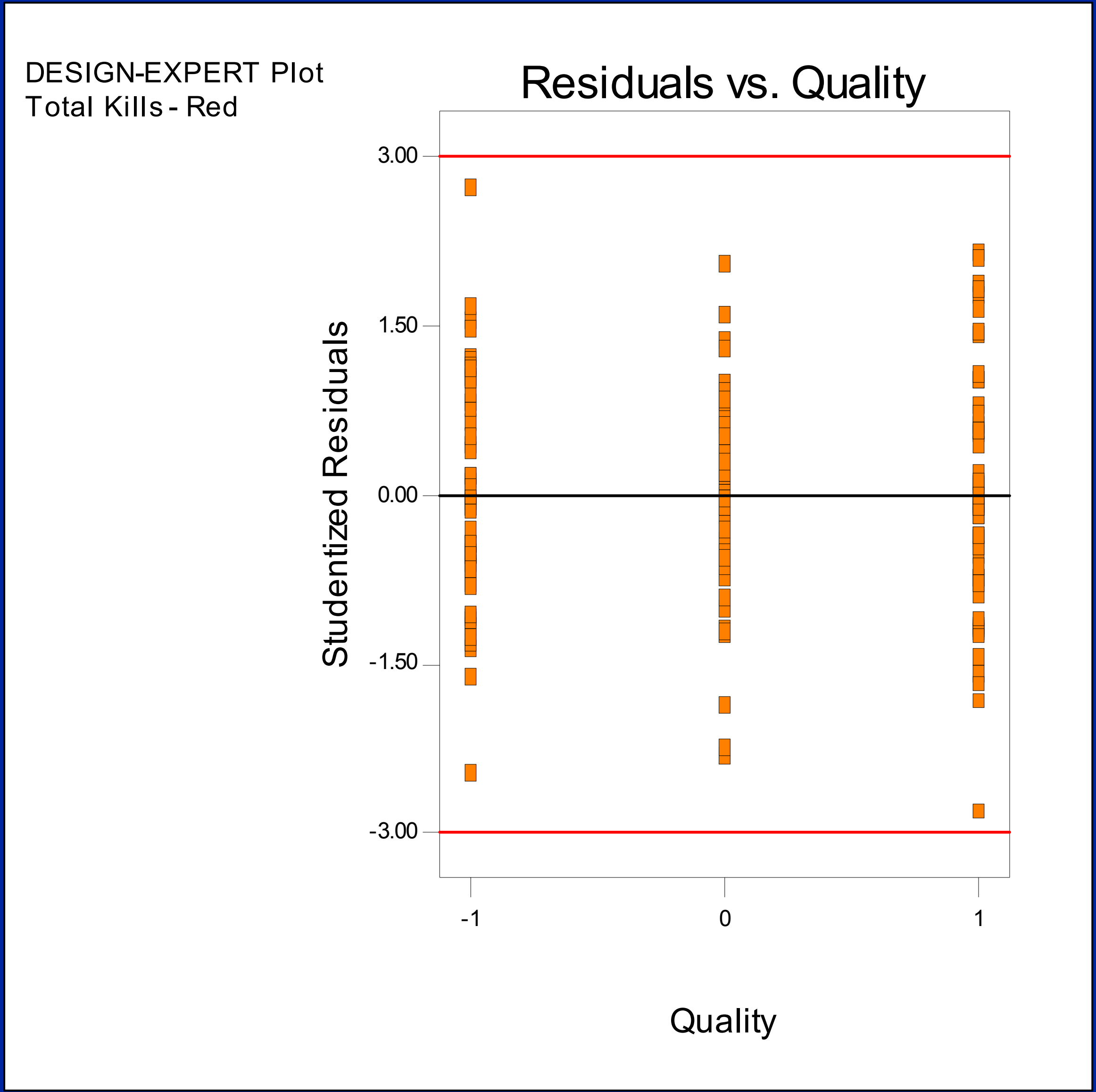
- Residual Analysis
  - Desired – no apparent pattern in the observed data
  - Actual – no pattern in the observed data
  - Results – OK



Residuals vs. Quality

DoE Tutorial

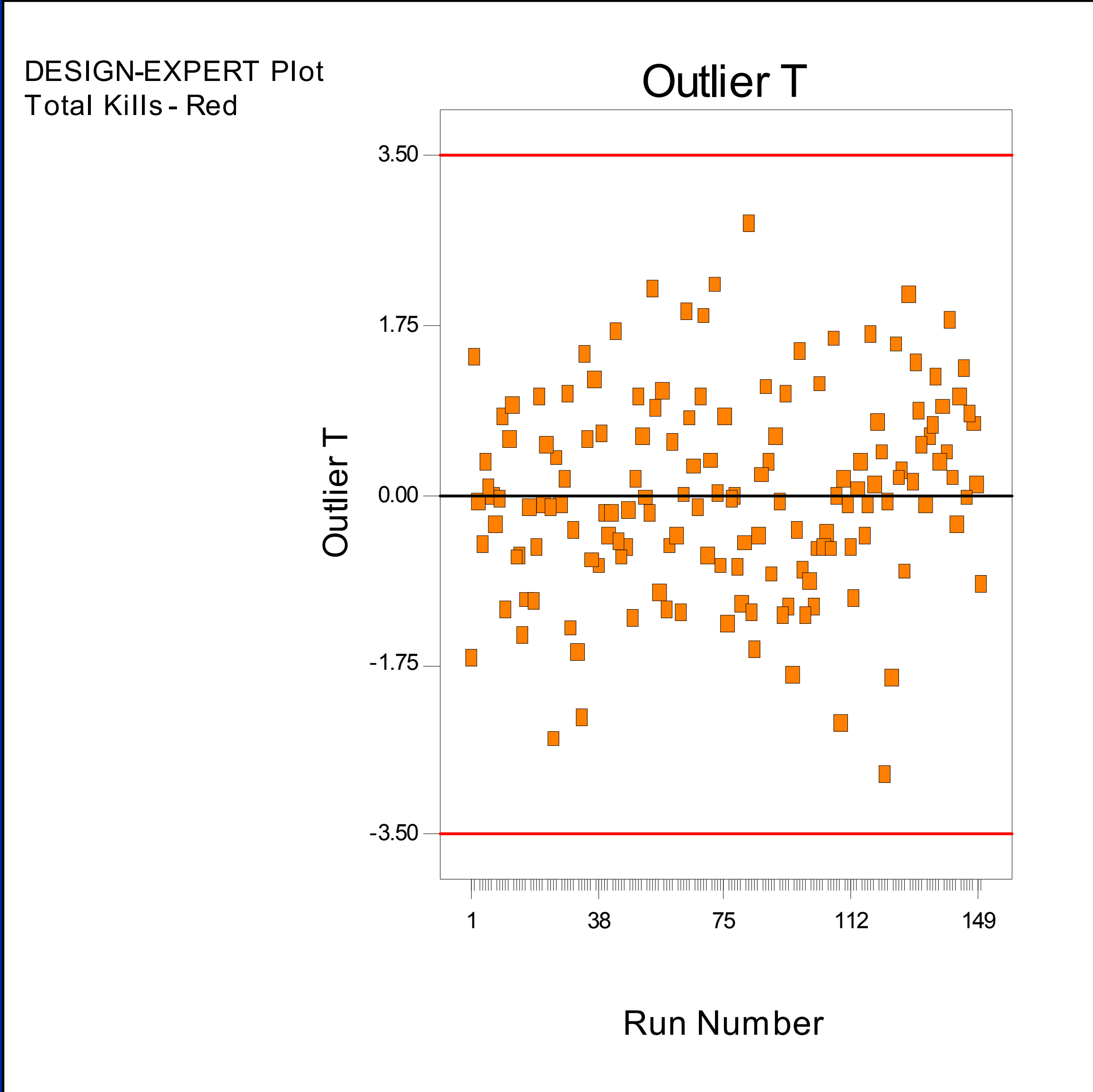
- Residual Analysis
  - Desired – no apparent pattern in the observed data
  - Actual – no pattern in the observed data
  - Results – OK



Outlier Investigation (1)

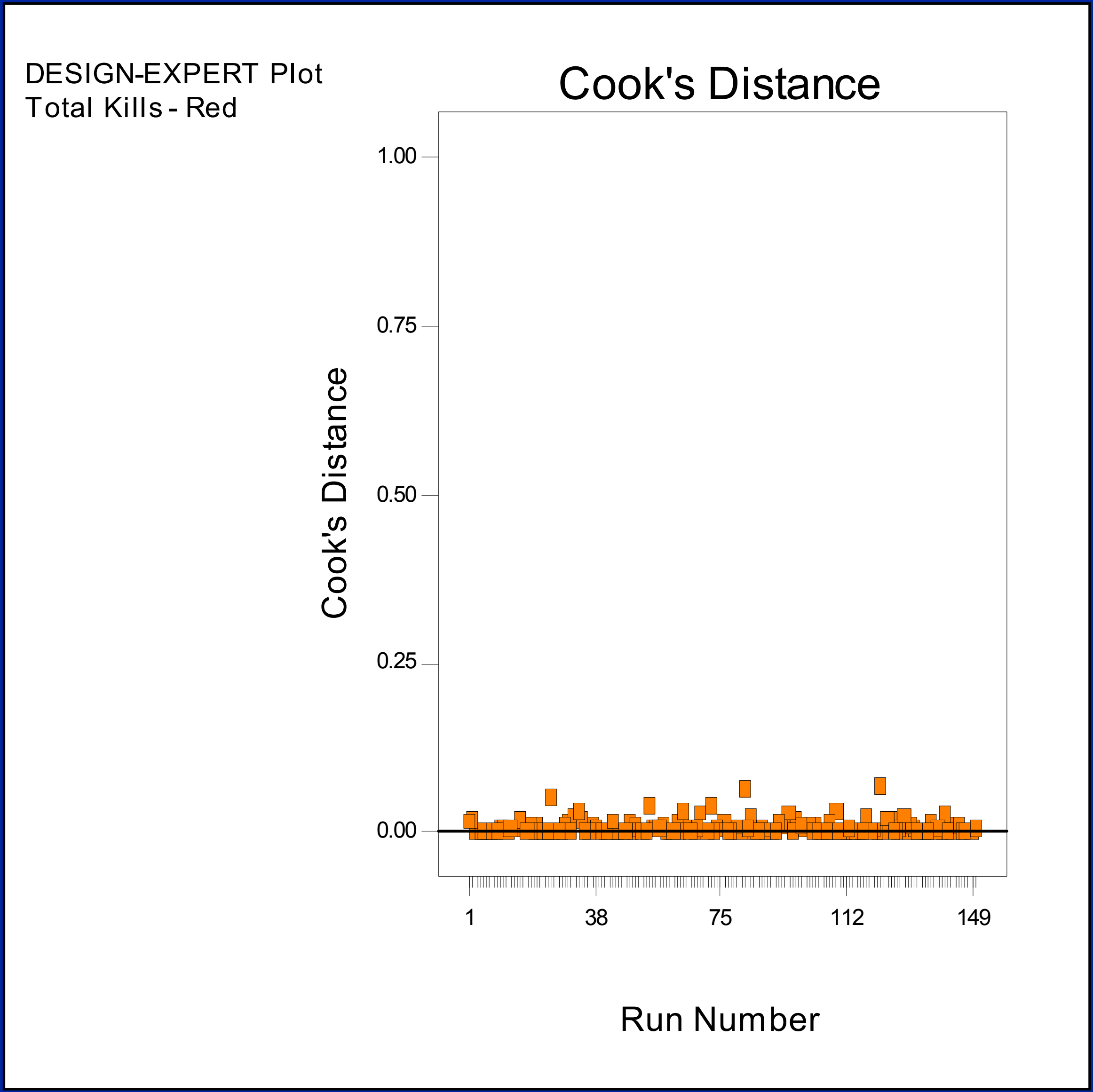
DoE Tutorial

- Outlier Analysis
  - Desired – no apparent pattern in the observed data
  - Actual – no pattern in the observed data
  - Results – OK



Outlier Investigation (2)

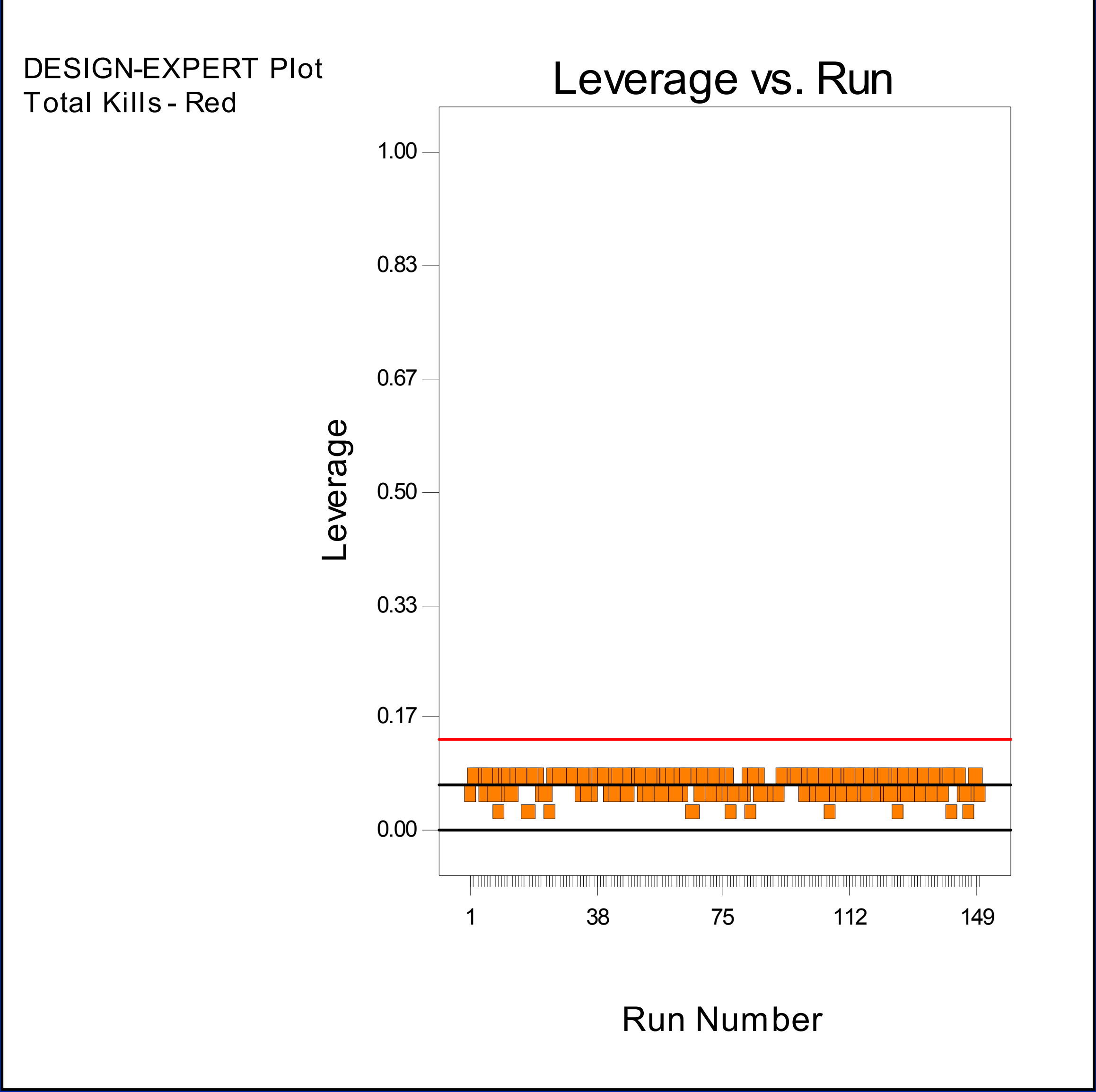
- Outlier Analysis
  - Desired – strong clustering near the zero point
  - Actual – strong clustering near the zero point
  - Results – OK



Outlier Investigation (3)

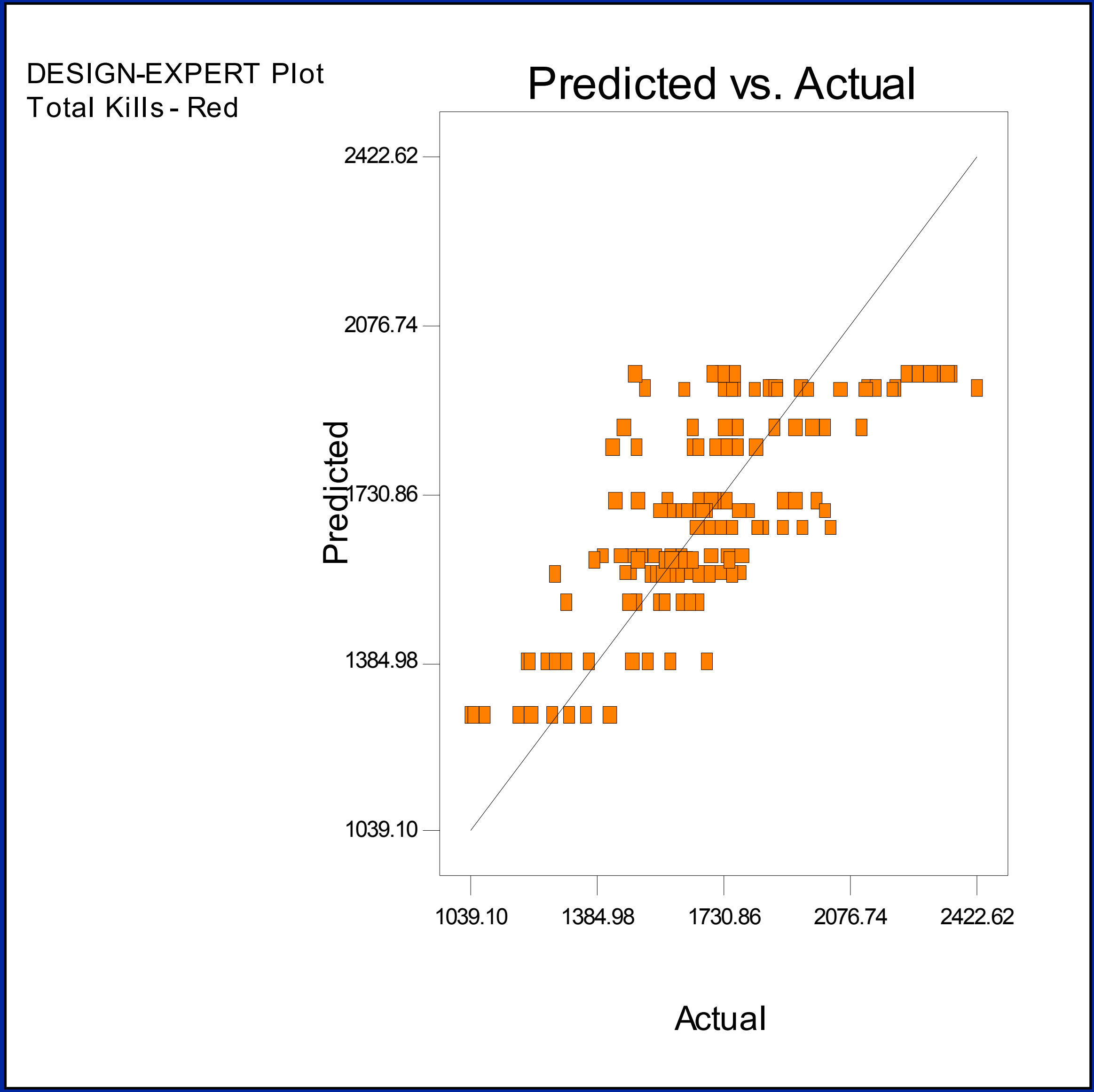
DoE Tutorial

- Outlier Analysis
  - Desired – strong clustering near the zero point
  - Actual – strong clustering near the zero point
  - Results – OK



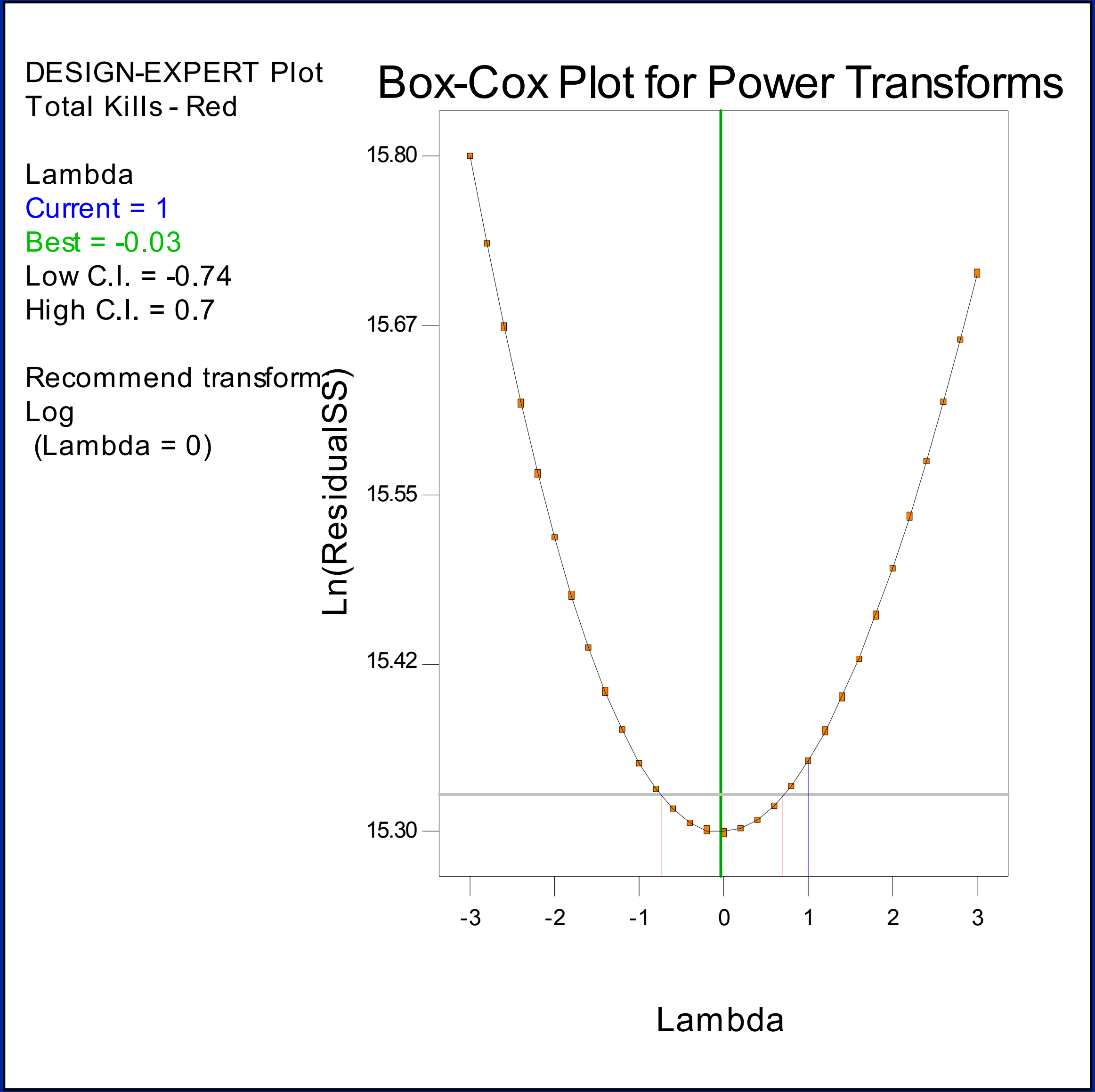
Predicted vs. Actual

- Outlier Analysis
  - Desired – no apparent pattern in the observed data
  - Actual – no pattern in the observed data
  - Results – OK



Power Transform

- Transform Analysis
  - Desired – no transform
  - Actual – Log transform recommended
  - Results – transform not pursued due to constraints:
    - Time available
    - Quality of data



Evaluation Results

- Model has statistical power:
  - For Type I error of 5%, Type II error is less than 0.1%.
- Diagnostics acceptable:
  - No problems based on residual analysis.
  - No problems based on outlier analysis.
  - Data transform suggested but not deemed essential for this task.

Interpretation and Analysis

- Review the results in terms of:
  - ANOVA Table
  - Perturbation plots
  - Single factor response
  - Interaction response
  - Contour plots
  - 3-D surface plots
  - Cube plot

ANOVA – Total Kills

DoE Tutorial

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	$\alpha = 0.05$
Model	6,024,287.18	9	669,365.24	20.2233	< 0.0001	*
T	3,235,042.05	1	3,235,042.05	97.739	< 0.0001	*
Q_T	659,344.82	1	659,344.82	19.9205	< 0.0001	*
Q_L	794,562.33	1	794,562.33	24.0058	< 0.0001	*
T^2	212,422.81	1	212,422.81	6.4178	0.0124	*
Q_T^2	558,906.79	1	558,906.79	16.886	< 0.0001	*
Q_L^2	356,637.37	1	356,637.37	10.7749	0.0013	*
TQ_T	17,652.64	1	17,652.64	0.5333	0.4664	
TQ_L	47,410.91	1	47,410.91	1.4324	0.2334	
Q_TQ_L	199,200.94	1	199,200.94	6.0184	0.0154	*
Residual	4,633,827.61	140	33,098.77			
Lack of Fit	660,829.65	5	132,165.93	4.4909	0.0008	*
Pure Error	3,972,997.96	135	29,429.61			
Cor Total	10,658,114.79	149				

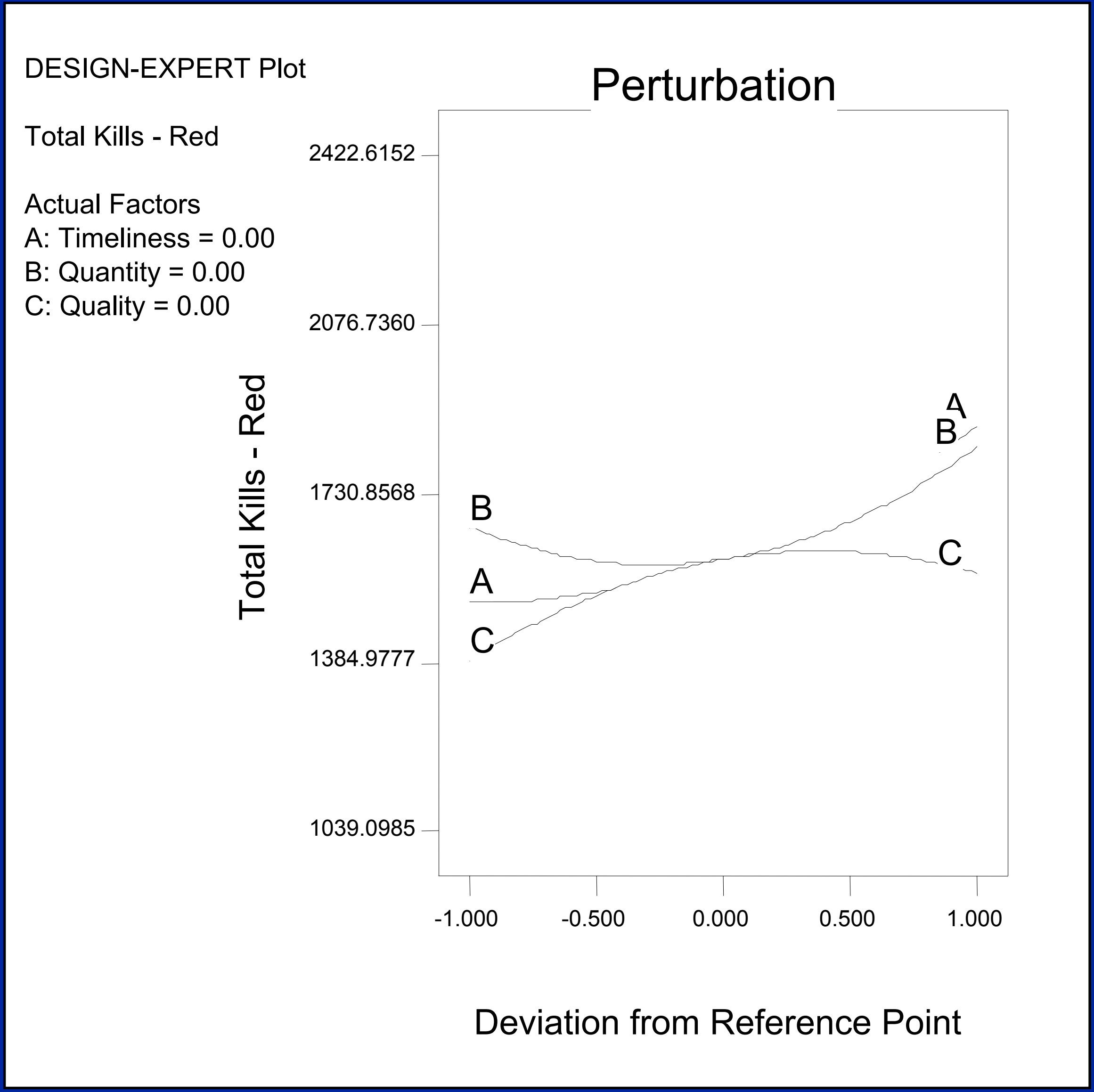
Fitted Model – Total Kills

$$CO = \left(\begin{array}{c} 1595.89 + 179.86T + 81.20Q_T + 89.14Q_L + \\ + 90.89T^2 + 147.43Q_T^2 - 117.77Q_L^2 \\ - 14.85TQ_T - 24.34TQ_L - 49.90Q_T Q_L \end{array} \right)$$

Perturbation Plot

DoE Tutorial

- Single Factor Analysis
  - Shows curvature for each factor at the Center point
  - Provides visual confirmation of the ANOVA statistics
  - “Opposing” shift for Quality (C) reflects value of the squared term in the fitted equation

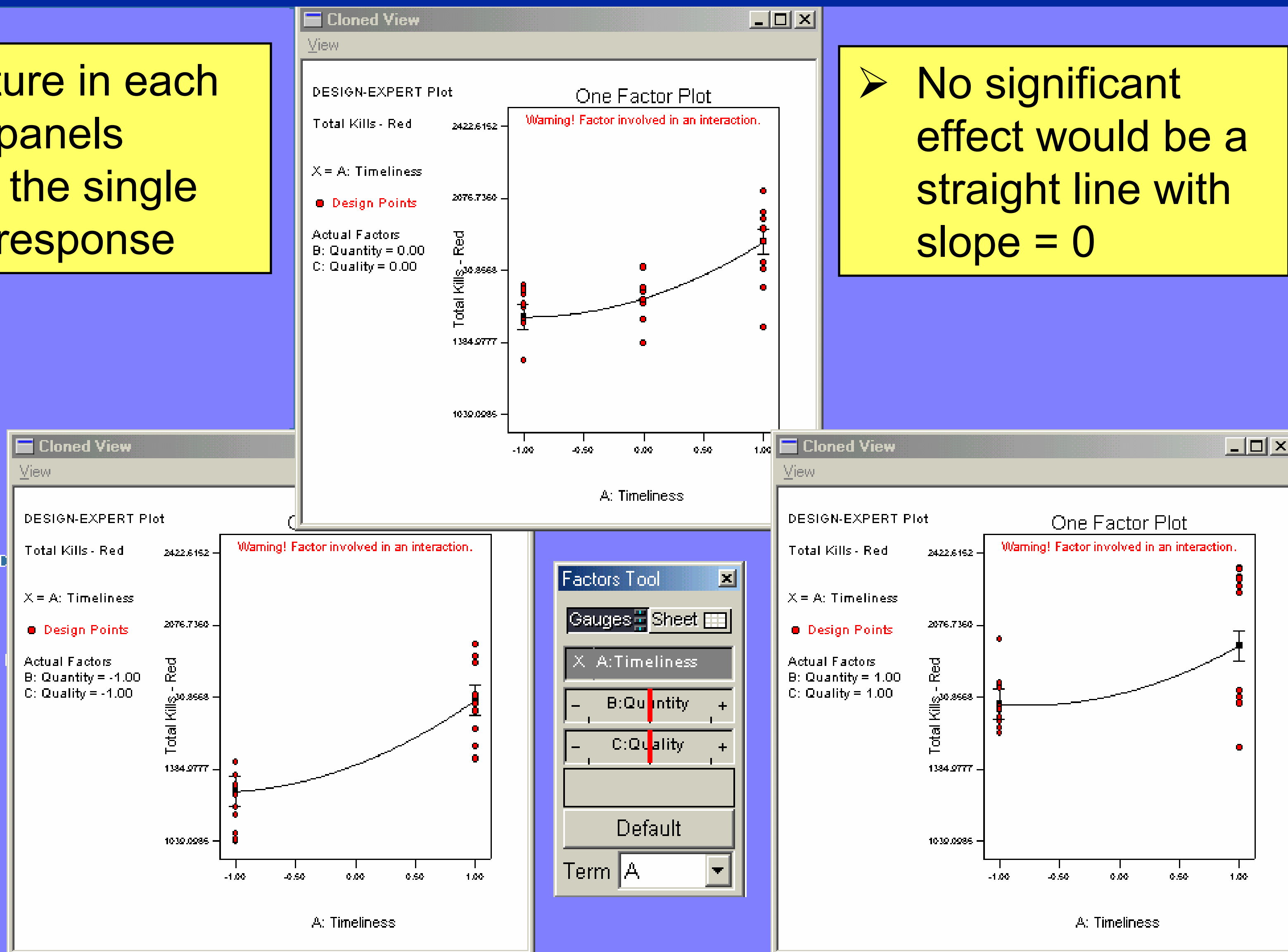


Single Factor - Timeliness

DoE Tutorial

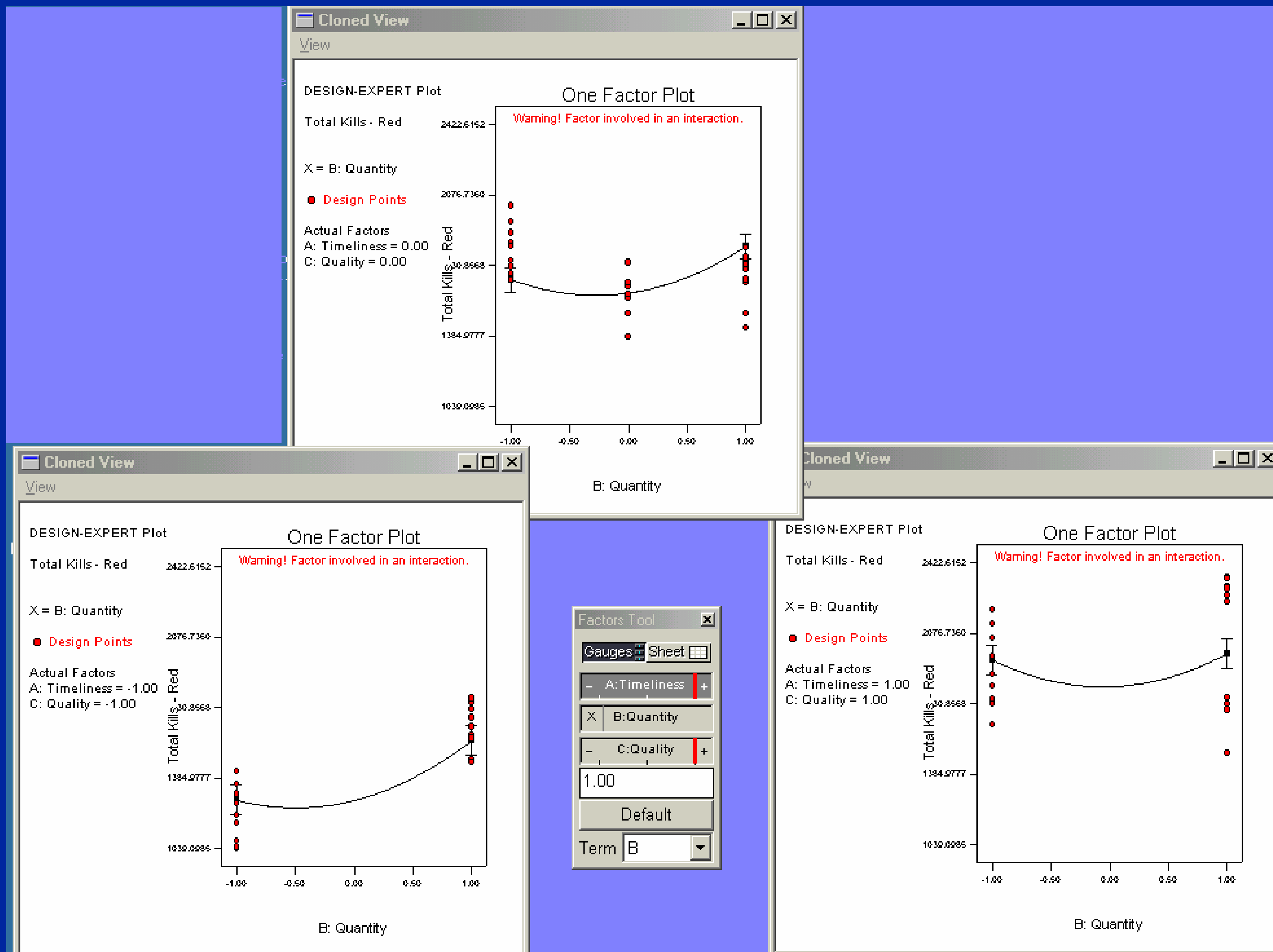
- Curvature in each of the panels shows the single factor response

- No significant effect would be a straight line with slope = 0



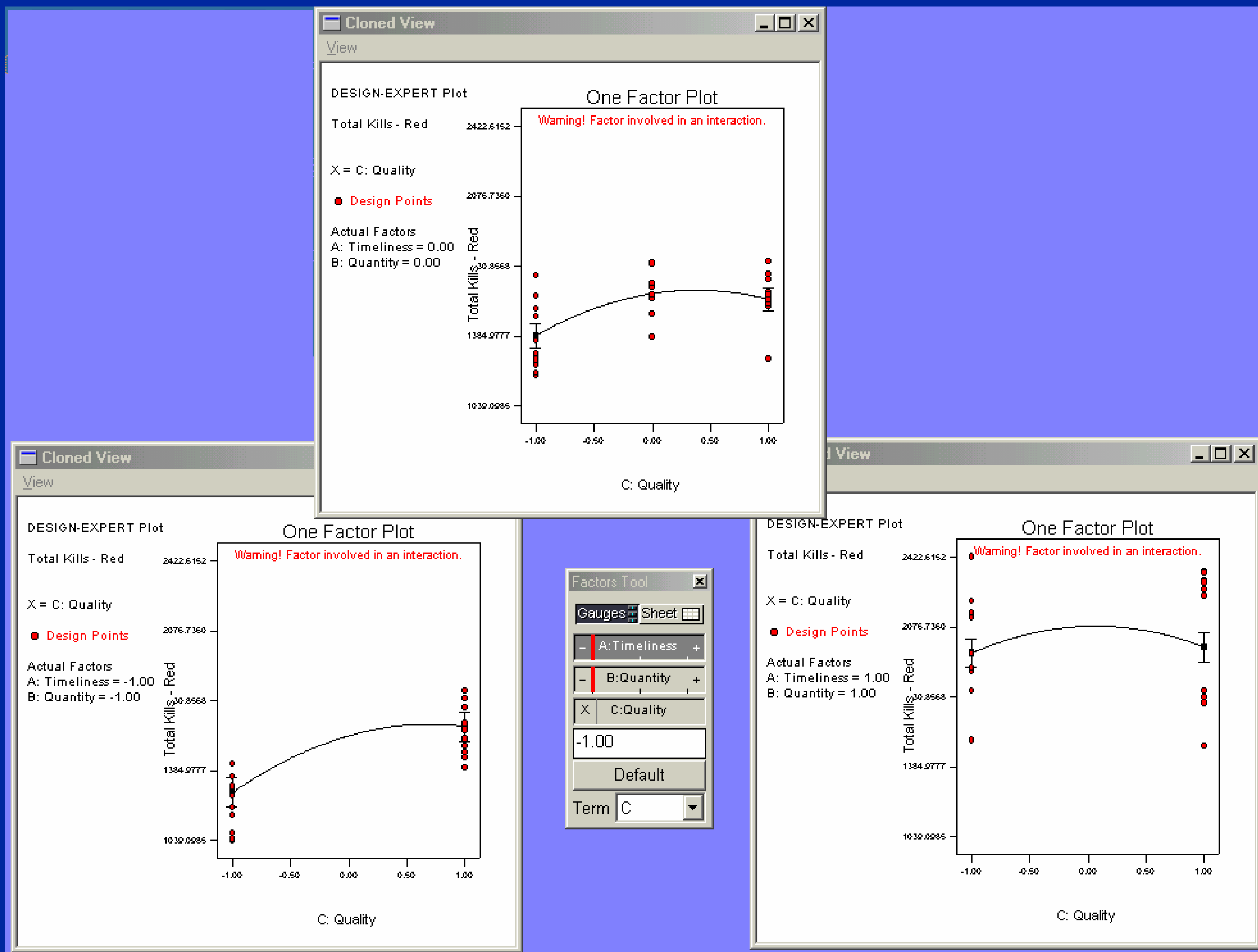
Single Factor - Quantity

DoE Tutorial



Single Factor - Quality

DoE Tutorial



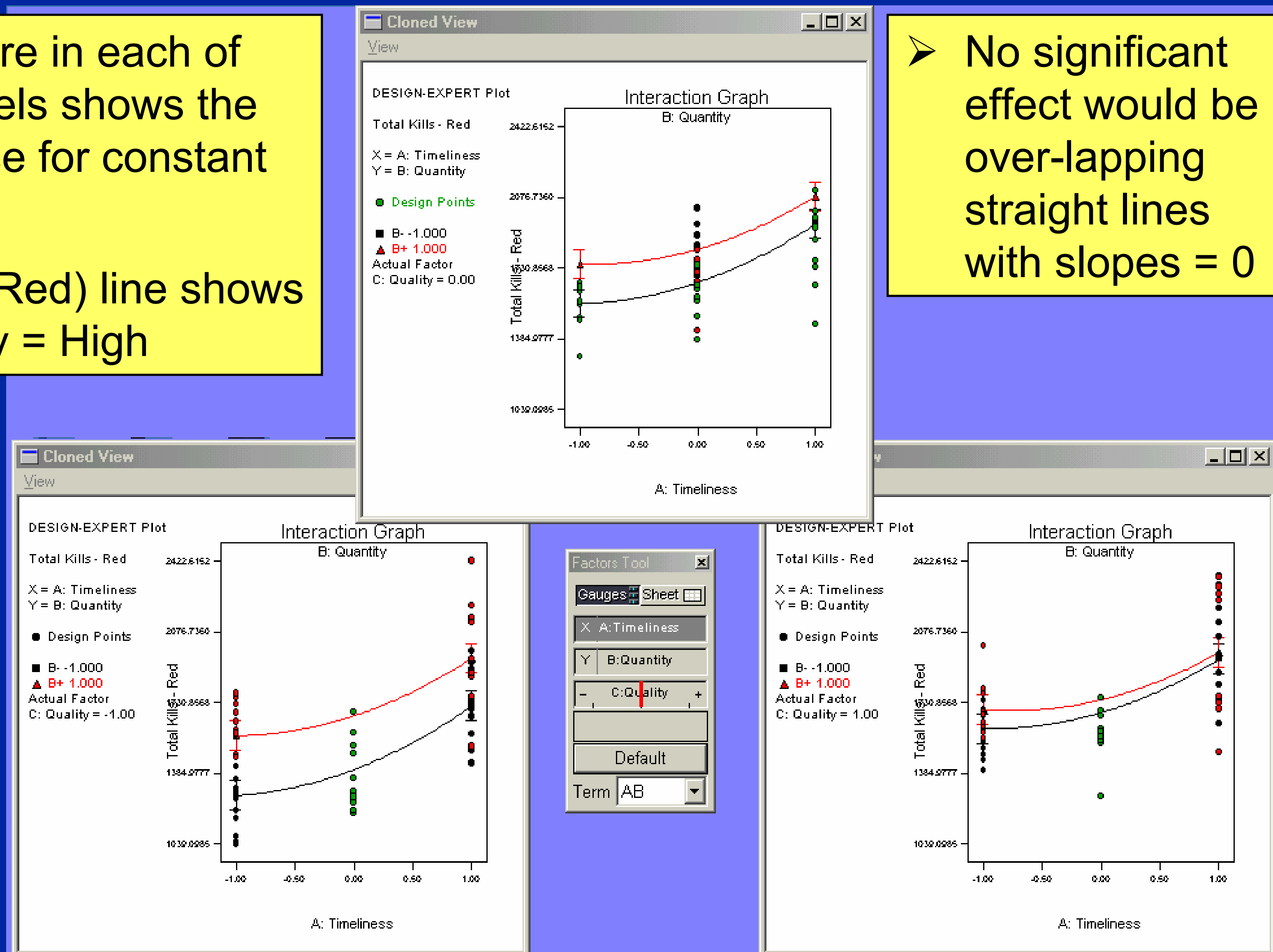
Lockheed Martin Center for Innovation

Interaction: T vs. Q_T

MORS

DoE Tutorial

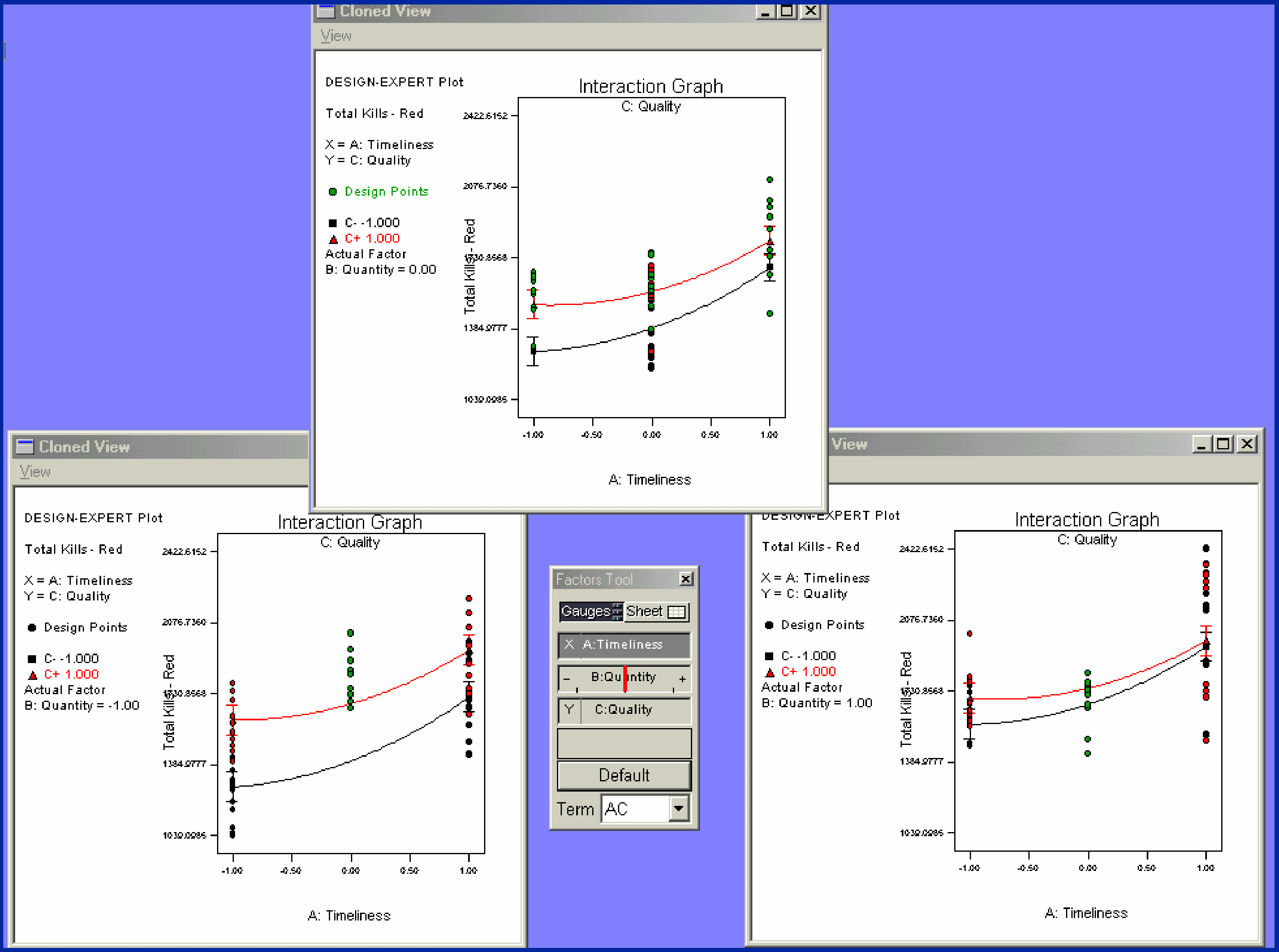
- Curvature in each of the panels shows the response for constant Quality
- Upper (Red) line shows Quantity = High



- No significant effect would be over-lapping straight lines with slopes = 0

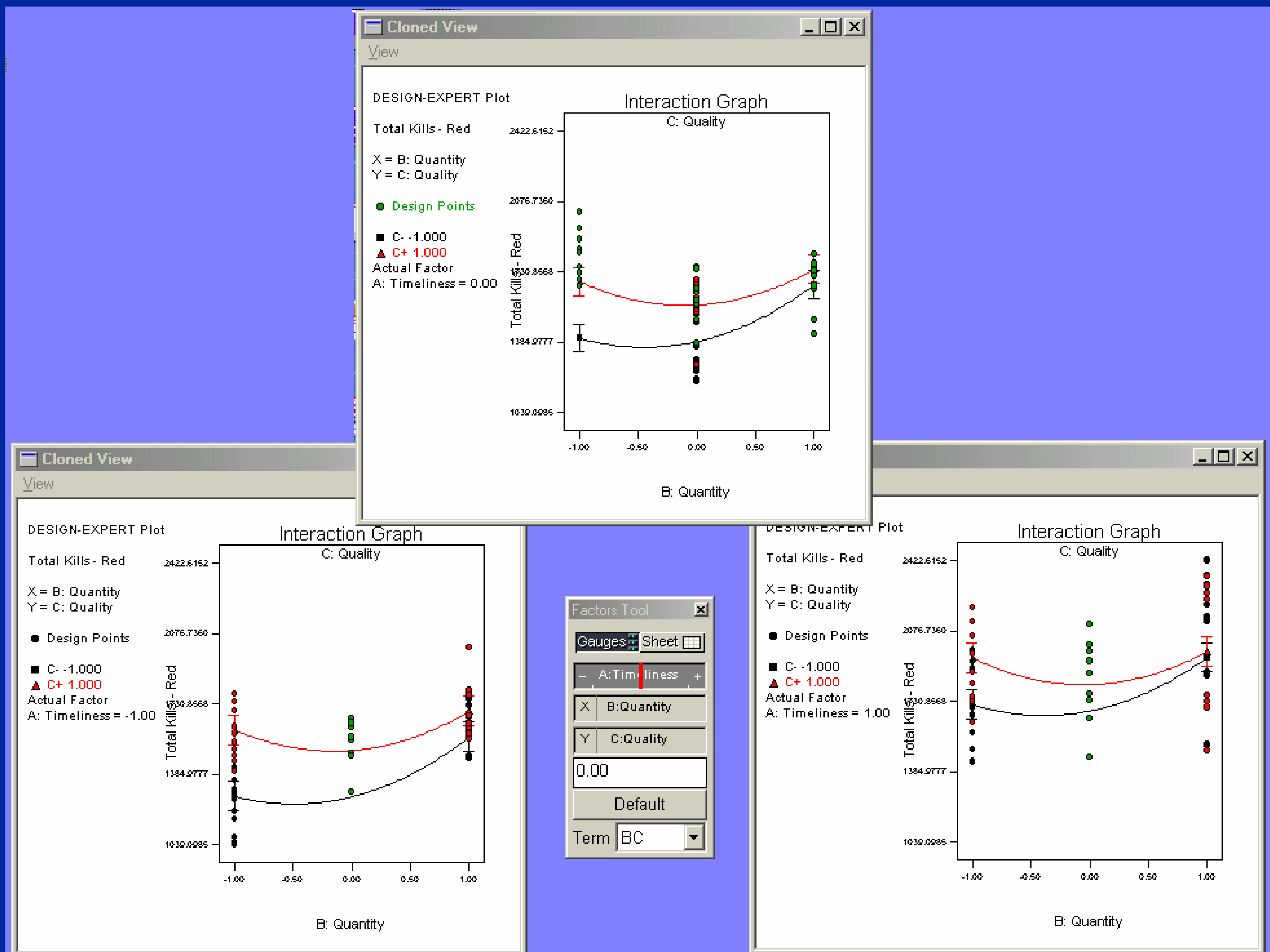
Interaction: T vs. Q_L

DoE Tutorial



Interaction: Q_T vs. Q_L

DoE Tutorial

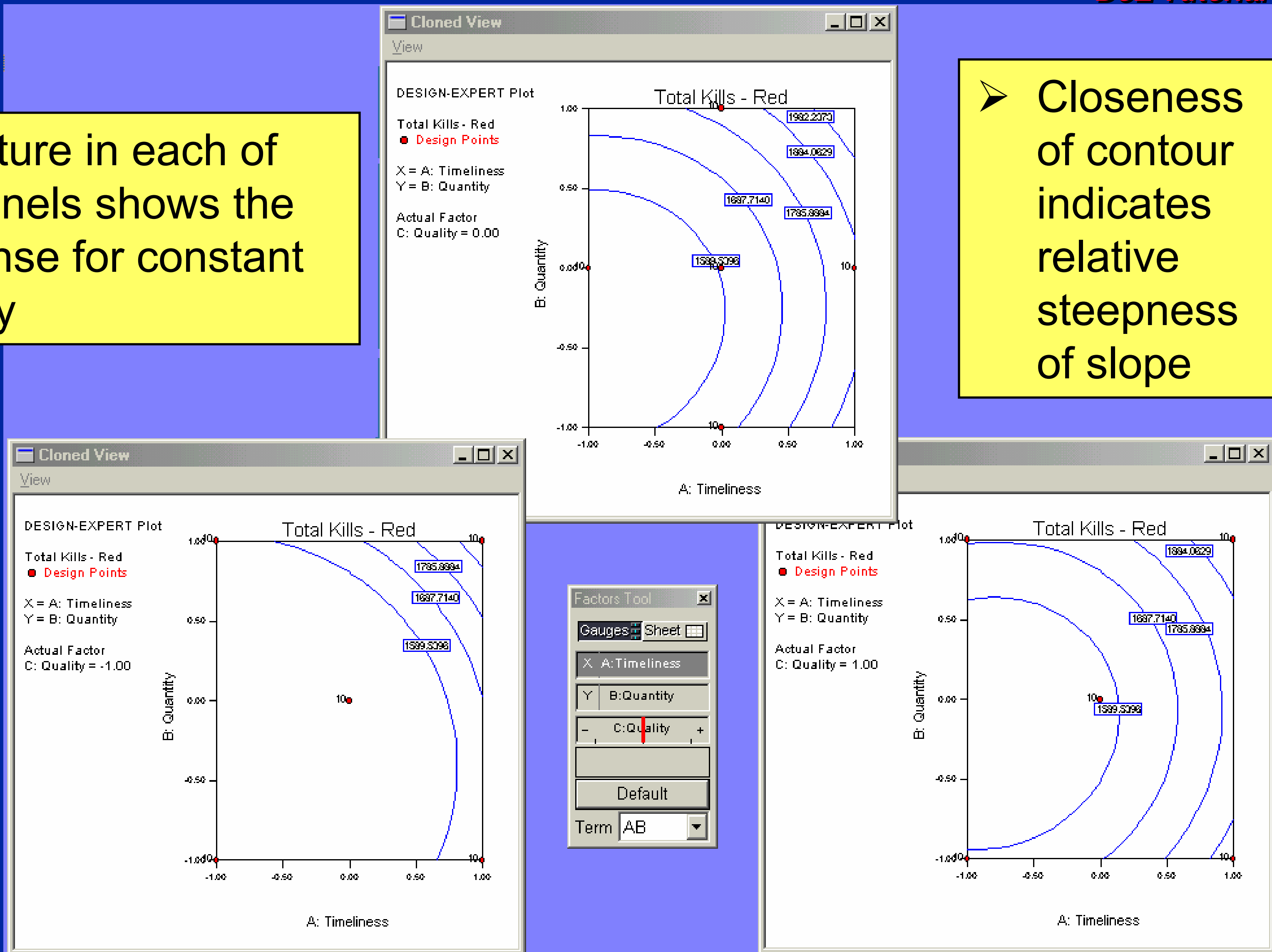


Lockheed Martin Center for Innovation

Contours: T vs. Q_T

DoE Tutorial

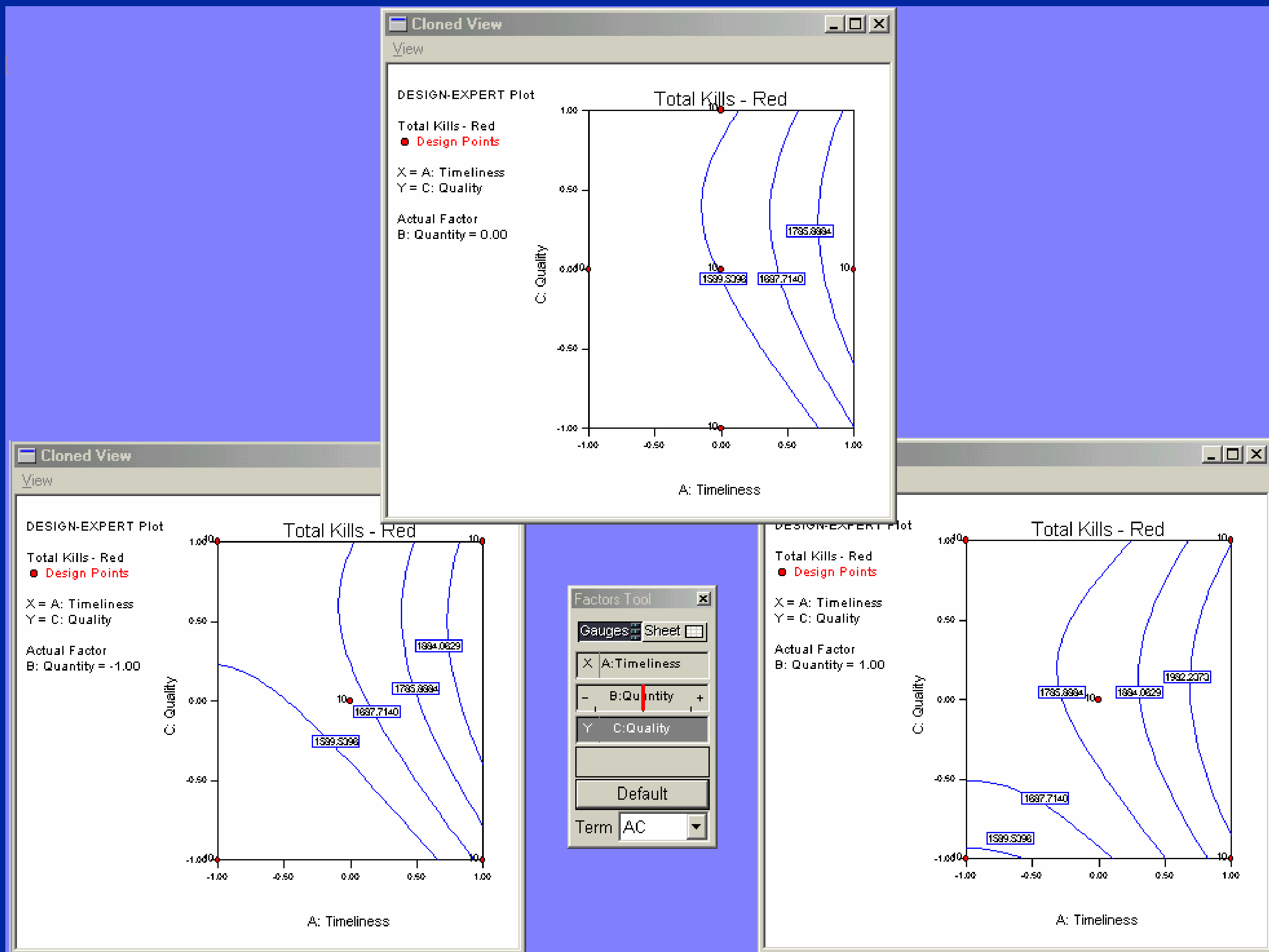
- Curvature in each of the panels shows the response for constant Quality



- Closeness of contour indicates relative steepness of slope

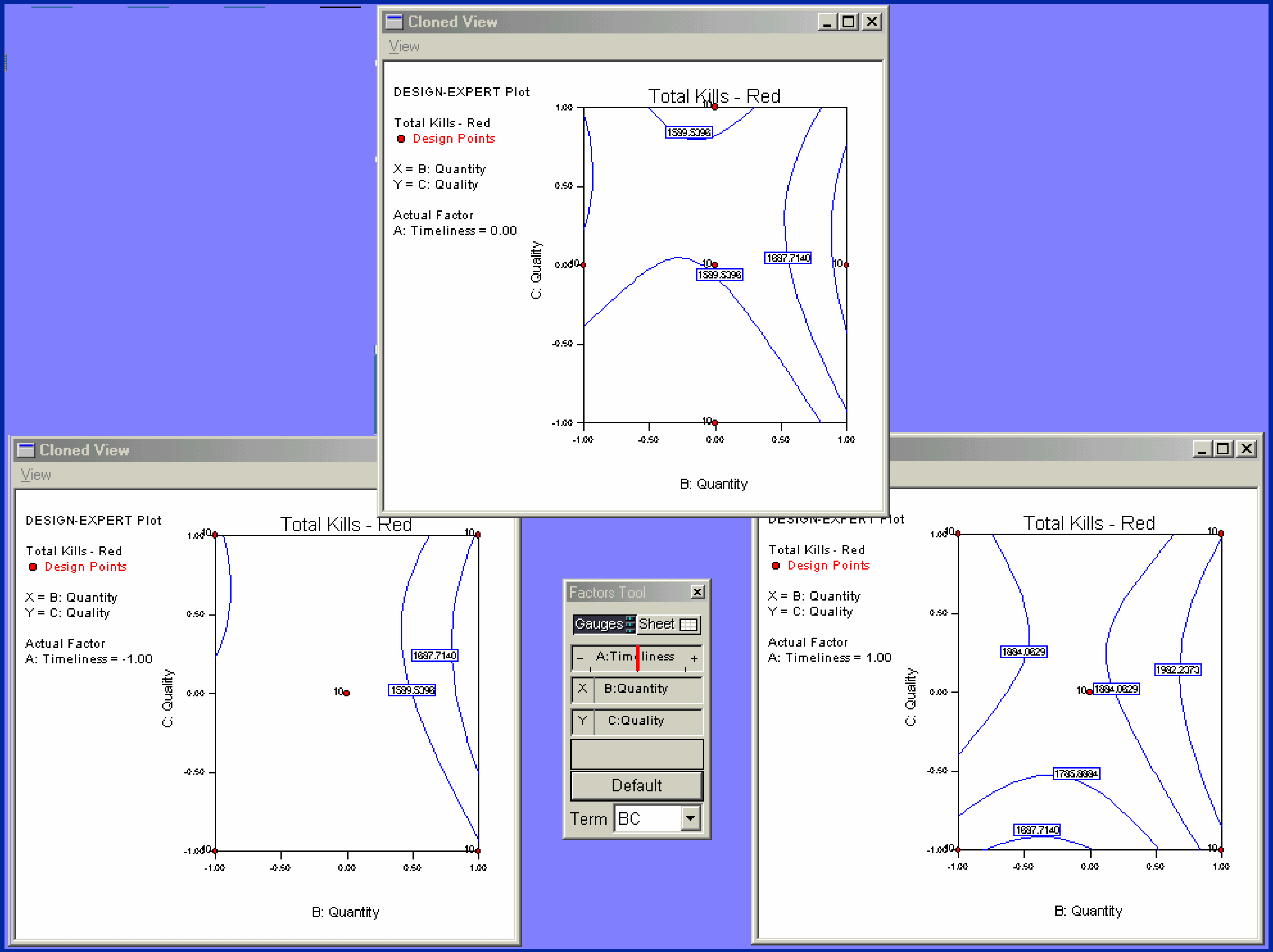
Contours: T vs. Q_L

DoE Tutorial



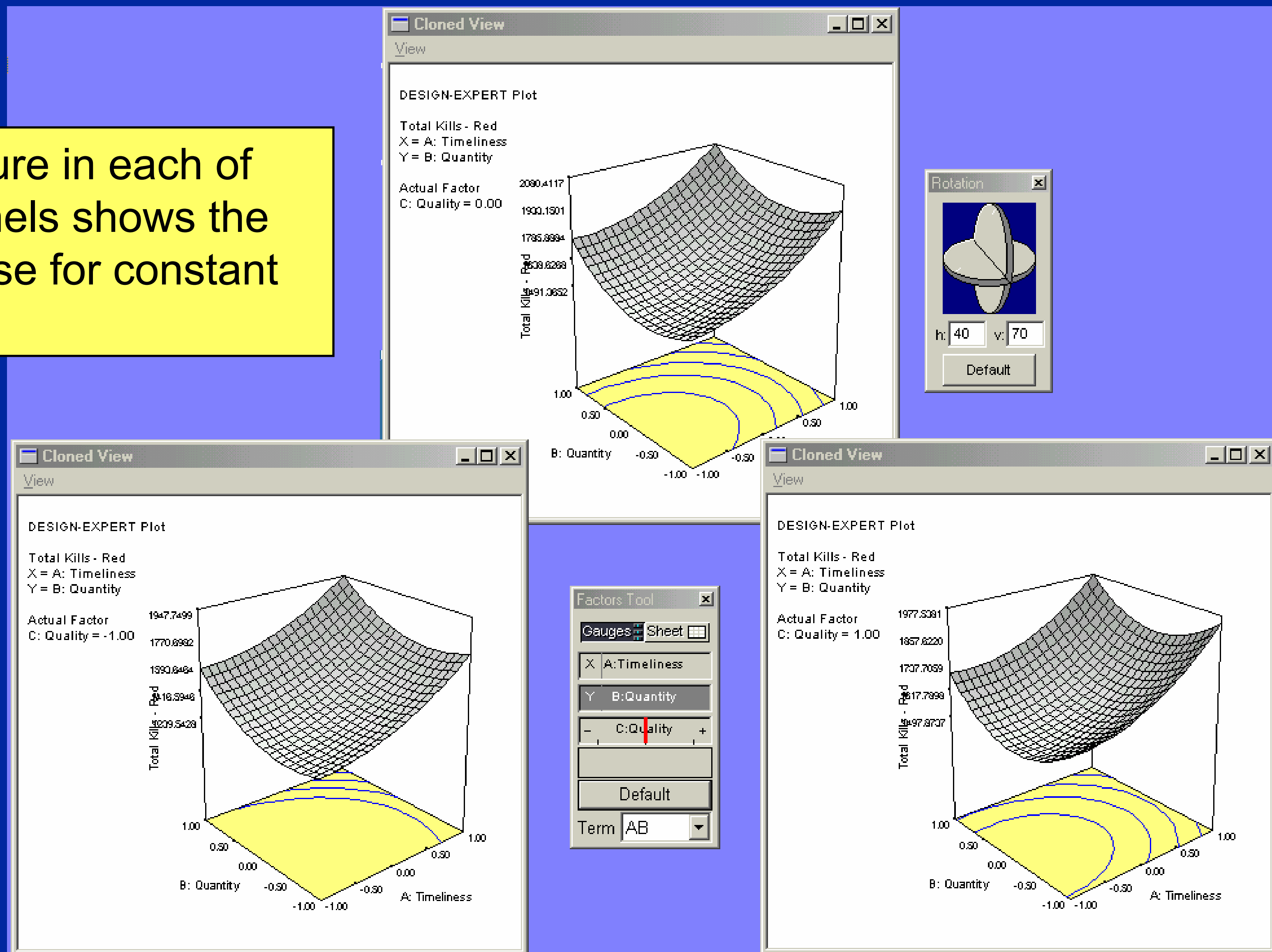
Contours: Q_T vs. Q_L

DoE Tutorial



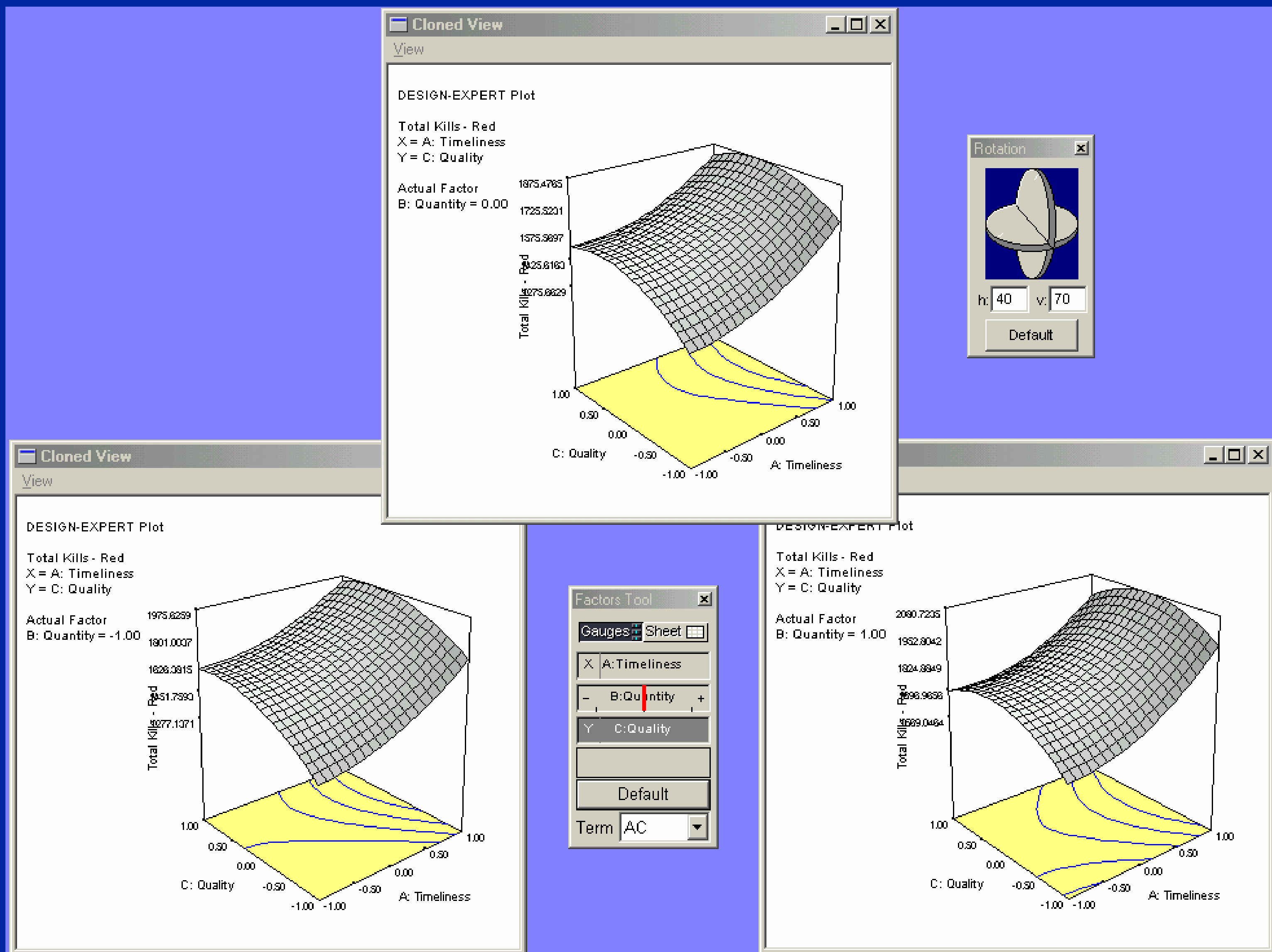
3-D Surface: T vs. Q_T

- Curvature in each of the panels shows the response for constant Quality



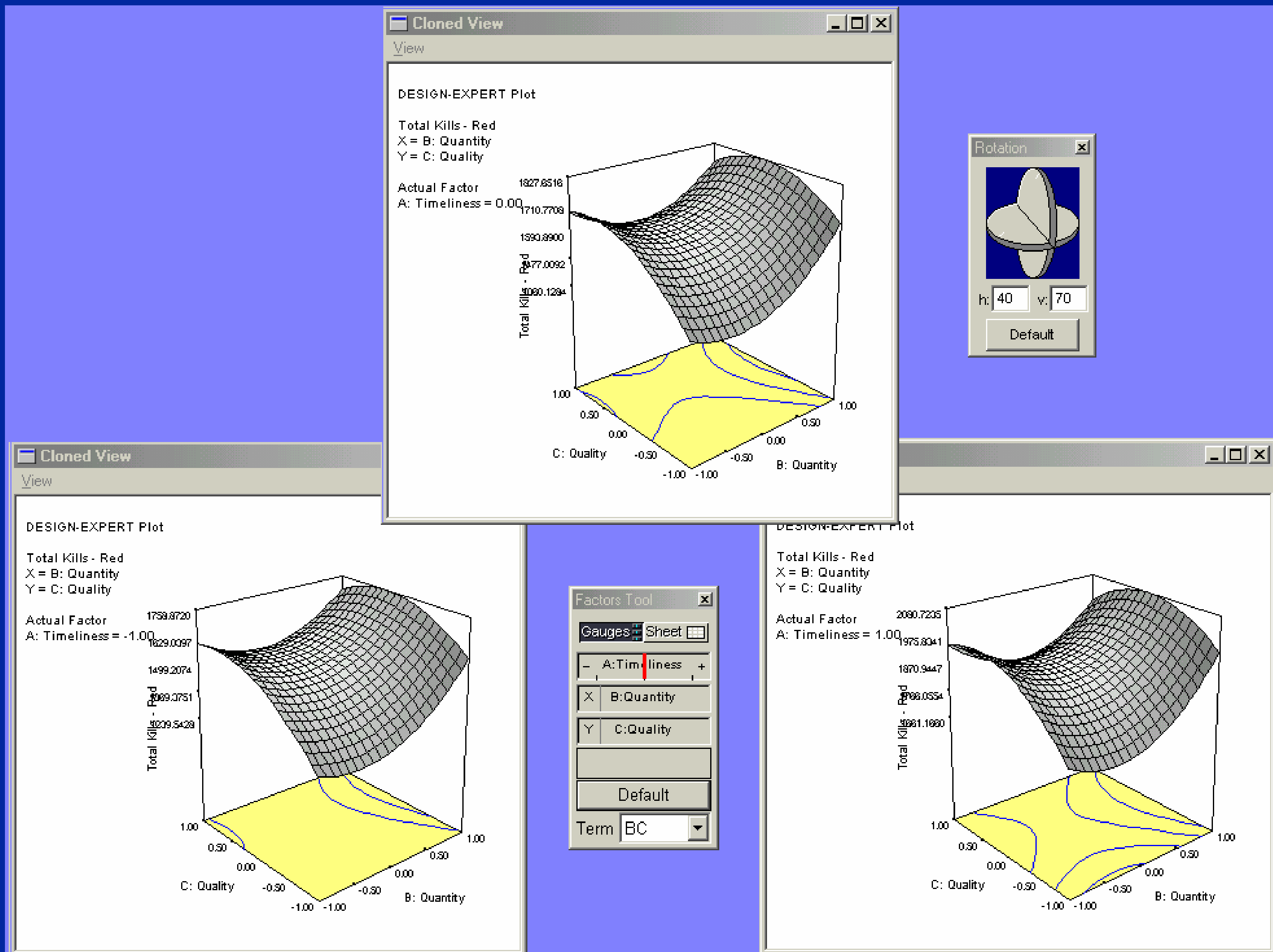
3-D Surface: T vs. Q_L

DoE Tutorial



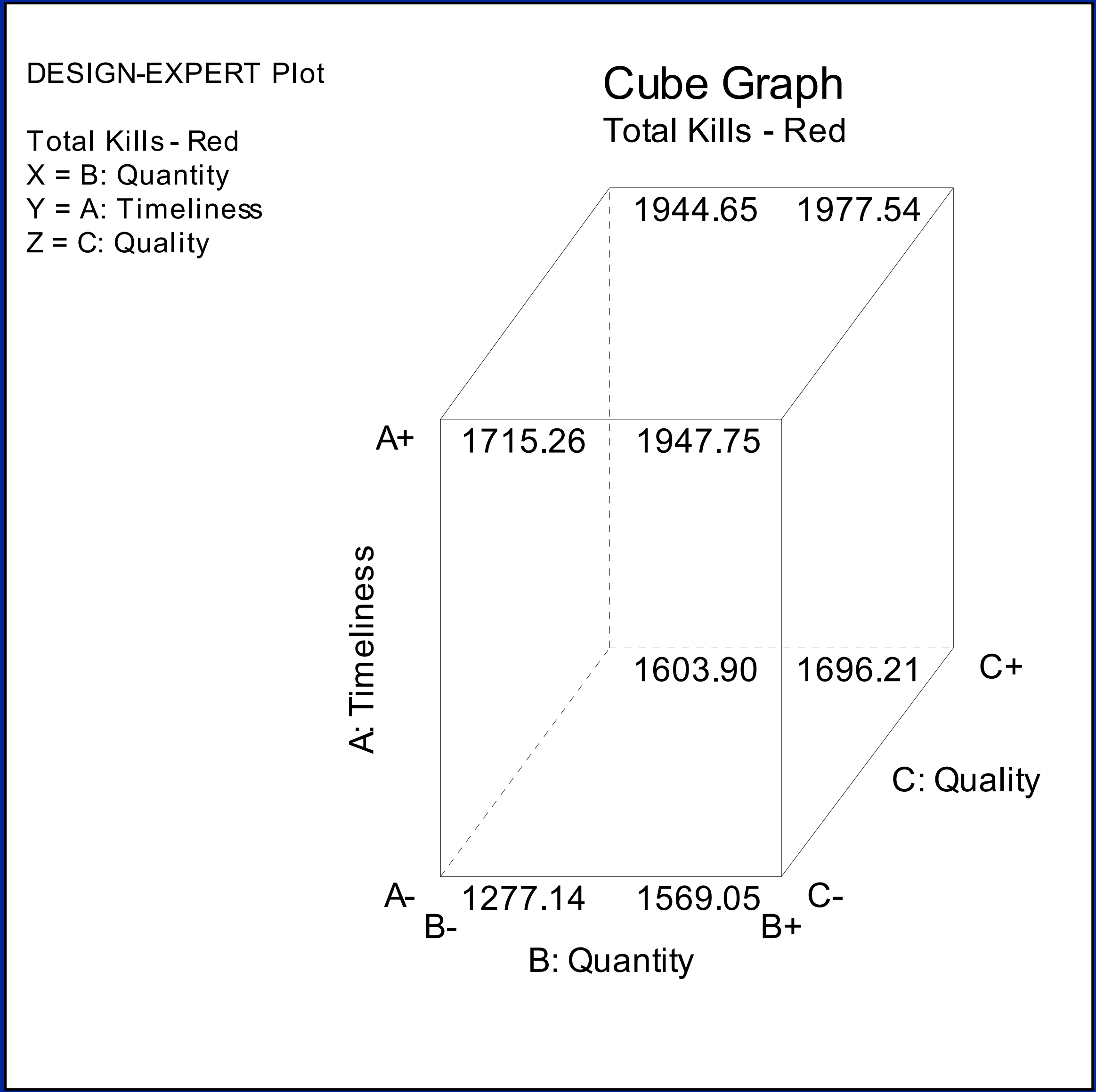
3-D Surface: Q_T vs. Q_L

DoE Tutorial



Design-Response Cube

- Response Analysis
  - Desired – Combat Outcome (Total Kills) increases as performance moves from degraded to enhanced
  - Actual – matches desired outcome
  - Results – model is sensitive to the 3 factors in the direction hypothesized



Significance Across Components

Factor	DF	IF	A2G	TOTAL
Model	*	*	*	*
T	*	*	*	*
Q_T	*		*	*
Q_L			*	*
T^2		*	*	*
Q_T^2	*	*	*	*
Q_L^2	*	*	*	*
TQ_T				
TQ_L				
Q_TQ_L	*			*

C3I Factor Sensitivity (β Coefficients)

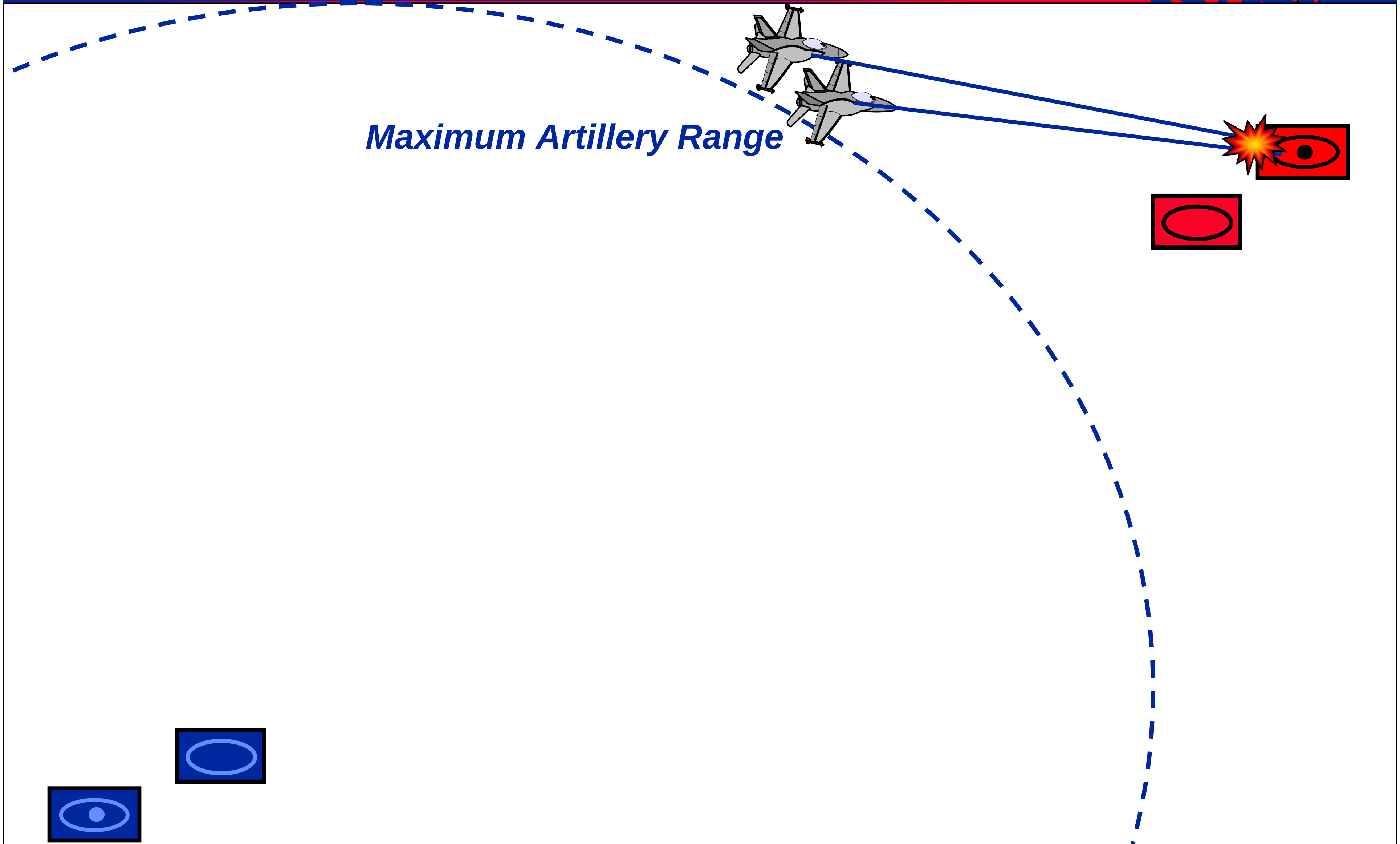
Lockheed Martin Center for Innovation

MORS

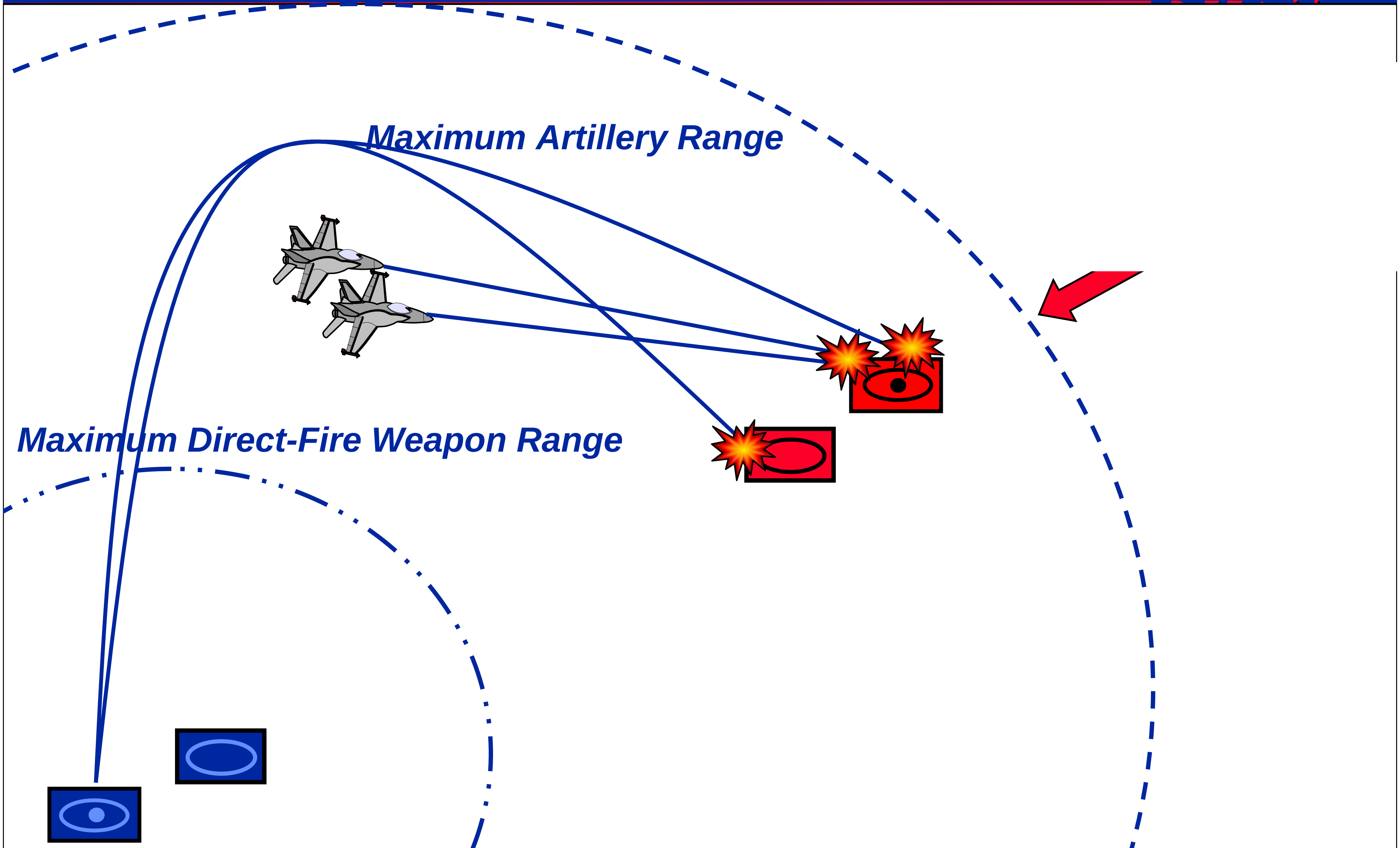
DoE Tutorial

Factor	DF	IF	A2G	TOTAL
β_0	286.13	141.07	1168.68	1595.89
T	-64.34	9.54	234.66	179.86
Q_T	-33.32	-0.29	114.81	81.20
Q_L	4.45	6.67	78.01	89.14
T^2	17.49	-48.30	121.70	90.89
Q_T^2	34.36	47.57	65.50	147.43
Q_L^2	-28.97	-20.69	-68.10	-117.77
TQ_T	8.30	7.21	-30.36	-14.85
TQ_L	9.42	-1.72	-32.05	-24.34
Q_TQ_L	-13.43	-4.90	-31.58	-49.90

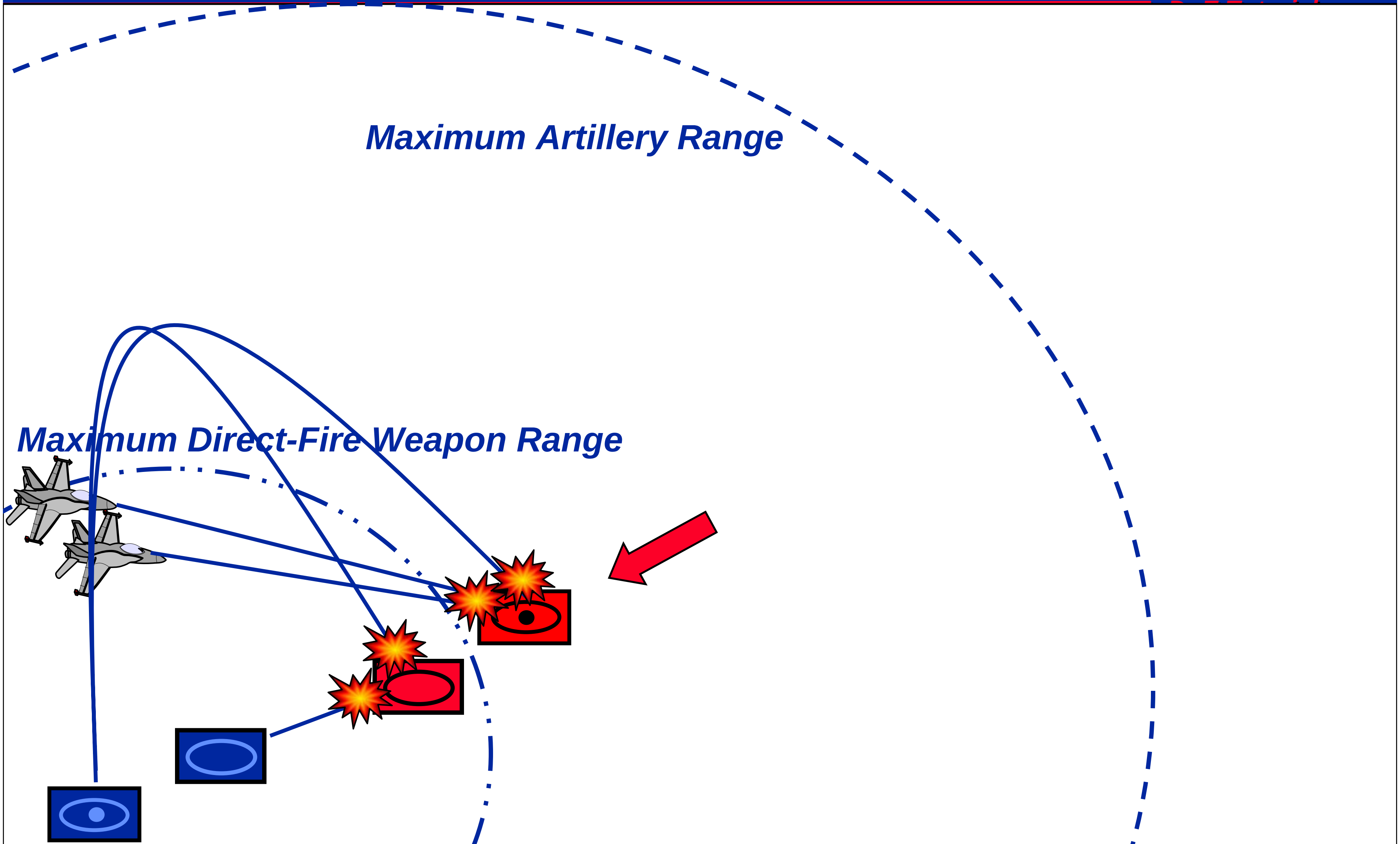
Battlefield Interactions



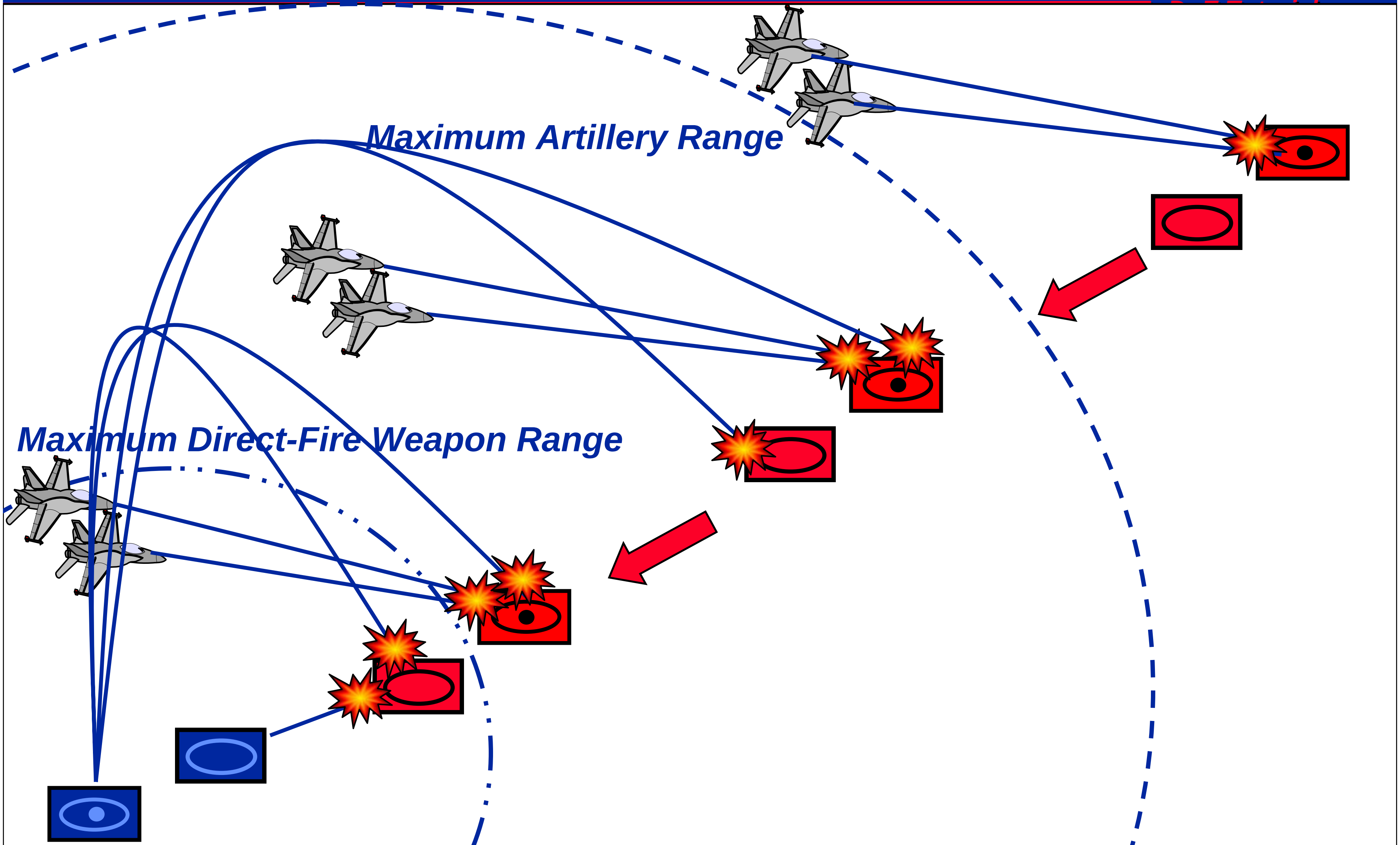
Battlefield Interactions



Battlefield Interactions



Battlefield Interactions



Conclusions

- The model is:
  - Sensitive to the C3I parameters of Timeliness, Quantity, and Quality of Information
- The Face-Centered CCD is:
  - Statistically Powerful
  - Robust
  - Capable of providing significant insights

Wrap-Up

- Topics Covered:
  - History from early days to Code of Best Practices
  - Types of experiments and why we do them
  - Strategies for experimentation
  - Basic comparison techniques
  - Analysis of Variance (ANOVA)
  - Complex variations of ANOVA
  - Importance of checking the diagnostic statistics
  - Practical military modeling example

Thank you for attending today.